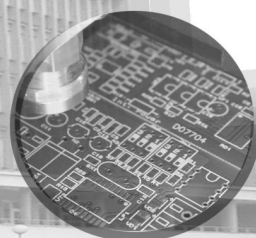# 11th Scientific Conference of Young Researchers

SCYR 2011

**Faculty of Electrical Engineering and Informatics
Technical University of Košice**

Proceedings from Conference

May 17th, 2011
Herľany, Slovakia

Sponsors

11th Scientific Conference of Young Researchers
of Faculty of Electrical Engineering and Informatics
Technical University of Košice

Proceedings from Conference

Program Committee of 11th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice

Chairman: Prof. Ing. Liberios Vokorokos, PhD.

Members: Prof. Ing. Roman Cimbala, PhD.
Prof. Ing. Dušan Kocur, CSc.
Prof. Ing. Iraida Kolcunová, PhD.
Prof. Ing. Ján Kollár, CSc.
Prof. Ing. Dobroslav Kováč, CSc.
Prof. Ing. Irena Kováčová, CSc.
Prof. Ing. Dušan Levický, CSc.
Prof. Ing. Ladislav Madarász, CSc.
Prof. Ing. Stanislav Marchevský, CSc.
Prof. RNDr. Valerie Novitzka, CSc.
Prof. Ing. Alena Pietriková, CSc.
Prof. Ing. RNDr. Ján Turán, DrSc.
Assoc. Prof. Ing. Ján Bača, CSc.
Assoc. Prof. Ing. Ľubomír Doboš, CSc.
Assoc. Prof. Ing. Želmíra Ferková, CSc.
Assoc. Prof. Ing. Zdeněk Havlice, CSc.
Assoc. Prof. Ing. Anna Jadlovská, CSc.
Assoc. Prof. Ing. Jozef Juhár, CSc.
Assoc. Prof. Ing. Marián Mach, CSc.
Assoc. Prof. Ing. Kristína Machová, CSc.
Assoc. Prof. Ing. Ján Paralič, PhD.
Assoc. Prof. Ing. Daniela Perduková, PhD.
Assoc. Prof. Ing. Jaroslav Porubän, PhD.
Assoc. Prof. Ing. Ladislav Samuelis, CSc.
Assoc. Prof. Ing. Branislav Sobota, CSc.
Assoc. Prof. Ing. Ján Šaliga, CSc.
Assoc. Prof. Ing. Jaroslava Žilková, PhD.

Organization Committee of 11th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice

Members: Ing. Martin Bačko
Ing. Mišél Batmend
Ing. Lýdia Dedinská
Ing. Michal Jurčišin
Mgr. Jana Petrillová
Ing. Emília Pietriková
Ing. Adela Tušanová
Mgr. Gabriela Vasziová
Ing. Eva Vozáriková

Contact address: Faculty of Electrical Engineering and Informatics
Technical University of Košice
Letná 9
042 00 Košice
Slovak Republic

Foreword

Dear Colleagues,

SCYR (Scientific Conference of Young Researchers) is a Scientific Event focused on exchange of information among young scientists from Faculty of Electrical Engineering and Informatics at Technical University of Košice - series of annual events that was founded in 2000. Since 2000 the conference has been hosted by FEI TUKE with rising technical level and unique multi-cultural atmosphere. The Eleventh Scientific Conference of Young Researchers (SCYR 2011), conference of Graduates and Young researchers, was held on 17th May 2011. The primary aims of the conference, to provide a forum for dissemination of information and scientific results relating to research and development activities at the Faculty of Electrical Engineering and Informatics has been achieved. 132 participants mostly by doctoral categories were active in the conference.

Faculty of Electrical Engineering and Informatics has a long tradition of students participating in skilled labor where they have to apply their theoretical knowledge. SCYR is opportunities for doctoral and graduating students use this event to train their scientific knowledge exchange. Nevertheless, the original goal to represent a forum for the exchange of information between young scientists from academic communities on topics related to their experimental and theoretical works in the very wide spread field of electronics, telecommunication, electrotechnics, computers and informatics, cybernetics and Artificial intelligence, electric power engineering, remained unchanged.

11th Scientific Conference of Young Researchers at Faculty of Electrical Engineering and Informatics Technical University of Košice (SCYR 2011) was organized in a beautiful village Herľany. The Conference was opened in the name of dean prof. Ing. Liberios Vokorokos, PhD. by the vicedean of faculty, doc. Ing. Roman Cimbala, PhD. In his introductory address he noted the importance of the Conference as a forum for exchange of information and a medium for broadening the scientific horizons of its participants and stressed the scientific and practical value of investigations being carried out by young researchers.

The program of conferences traditionally includes two parallel sessions (both consist of oral and poster part):

- Electrical & Electronics Engineering

- Informatics & Telecommunications

with 110 technical papers dealing with research results obtained mainly in university environment. This day was filled with a lot of interesting scientific discussions among the junior researchers and graduate students, and the representatives of the Faculty of Electrical Engineering and Informatics. This Scientific Network included various research problems and education, communication between young scientists and students, between students and professors. Conference was also a platform for student exchange and a potential starting point for scientific cooperation. The results presented in papers demonstrated that the investigations being conducted by young scientists are making a valuable contribution to the fulfillment of the tasks set for science and technology at Faculty of Electrical Engineering and Informatics at Technical University of Košice.

We want to thank all participants for contributing to these proceedings with their high quality manuscripts. We hope that conference constitutes a platform for a continual dialogue among young scientists.

It is our pleasure and honor to express our gratitude to our sponsors and to all friends, colleagues and committee members who contributed with their ideas, discussions, and sedulous hard work to the success of this event. We also want to thank our session chairs for their co-operation and dedication throughout the whole conference.

Finally, we want to thank all the attendees of the conference for fruitful discussions and a pleasant stay in our event.

<div align="right">

Liberios VOKOROKOS
dean of FEI TUKE

</div>

May 17$^{\text{th}}$ 2011, Herľany

Contents

Section: Electrical & Electronics Engineering

Section: Informatics & Telecommunications

Section:
Electrical & Electronics Engineering

A Novel Concept of Photovoltaic System

Ján Perduľak

Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

jan.perdulak@tuke.sk

Abstract — **This article introduce a new concept of photovoltaic system. This new concept allows effective utilize of energy from photovoltaic solar cell. The effective utilize of energy is ensured by proper changes of turns ratio. The core of proposed photovoltaic system is full-bridge converter which ensures the constant voltage on the secondary side of transformer. The photovoltaic system is proposed for 12 V systems. The output energy is stored in battery. The theoretical assumes are verified by simulation results.**

Keywords — **photovoltaic system, full-bridge converter, turns ratio.**

I. INTRODUCTION

There are more and more increasing demand for the energy stocks and energy self-sufficient of the countries in these days. The logical consequence of this situation is increasing prices of non renewable sources of energy.

Using the renewable sources of energy can solve this situation. On the present the photovoltaic (PV) has a dominant position. Photovoltaic is the direct conversion of light into electricity in form of direct current electricity. Usually the DC/DC converters are use to convert this direct electrical power from one level to another.

There are many types of materials which are use to make of the PV modules. The main problem of these materials is low conversion efficiency. The conversion efficiency is from 5% (a-Si) to 25% - 30% (GaAs) [1]. To-days efficiency of the soft switching DC/DC converters is very well. It is moving around the 97%. But on the other hand the energy efficiency is not good in comparison with abovementioned converter efficiency.

This paper present the new concept of PV system with possibility of utilizes the full range of energy supplied from the solar cell. The core of proposed converter is full-bridge (FB) converter which ensures the constant voltage U_S on the secondary side of transformer. This constant voltage is ensure by tapped voltage transformer on the secondary side. The effective utilize of PV energy in full range of input voltages U_{PV} is ensures by switching of particular tapped windings.

II. THE PROPOSED CONCEPT OF PV SYSTEM

The main idea of proposed concept is in changing of turns ratio considering in changing of irradiance or more precisely solar cell output voltage. This solar cell voltage is sensing by control system and consequently the new turns ratio is set up.

This change of turns ratio ensure that the voltage on the secondary side of transformer U_S is constant Fig.1.
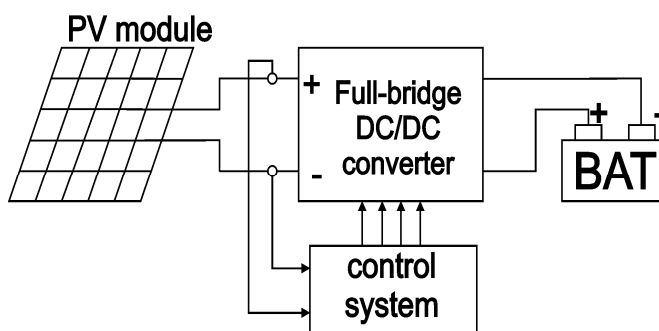
Fig. 1. The proposed concept of PV system.

Fig.2 shows the topology of FB converter. The FB converter consists of full-bridge inverter (INV), transformer (TR) with main and tapped windings on the secondary side, bridge rectifier (REC) and LC filter (F).
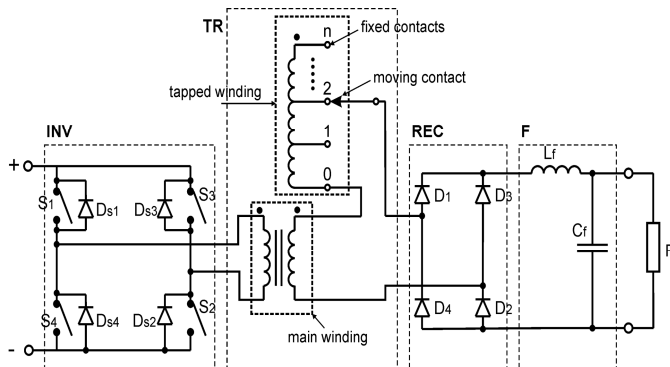
Fig. 2. The topology of FB converter.

III. PRINCIPLE OF OPERATION

Principle of operation of FB converter is well known. The DC voltage of solar cell is converted by inverter (INV) to the AC voltage which is consequently transformed from one level to another by transformer (TR). The secondary voltage of transformer is consequently rectified (convert to the DC voltage again) by using bridge rectified situated on the output of transformer (TR). Fig.3. shows the theoretical voltage and current waveforms of FB converter. In this case the diagonally opposite switches (S1 and S2, or S3 and S4) are turned on and off simultaneously.
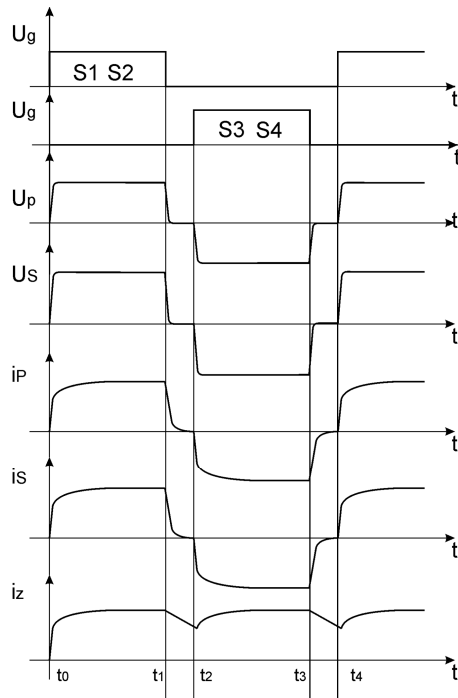
Fig. 3. The theoretical waveforms of FB converter.

IV. OPERATION ANALYSIS OF FB CONVERTER

The FB converter with the fixed value of turns ratio can optimally works (of course if it is design for particular system) only in limited range of input voltages. The proposed FB converter or more precisely PV system is design for optimally works in wide range of input voltage U_{PV}. The PV system is design for the 12 V system. The required secondary voltage U_S is 14V. The output energy of PV system is stored in battery.

The FB converter must be able to works in wide range of input voltages form solar cell. There is a demand for minimum input voltage U_{PV} (that means the worst case) in which the FB converter must be able to optimally work. The minimum input voltage was set up on 4V that means the there is request of 9 different turns ratio (9 tapped windings). This different turns ratio are determine form (1):

$$p_k = \frac{U_{PVk}}{U_S}, k = \sum_{l=1}^{9} l \qquad (1)$$

TABLE I
THE VALUES OF TURNS RATIO FOR DIFFERENT VALUES
OF INPUT VOLTAGE U_{PV}

k	U_{PV} [V]	p	U_S [V]
1	4	0,285	
2	5	0,357	
3	6	0,428	
4	7	0,5	
5	8	0,571	14
6	9	0,642	
7	10	0,714	
8	11	0,785	

| 9 | 12 | 0,857 | |

The inductance of primary winding L_P of transformer is set up on the constant value L_P = 100 mH. The values of inductance of secondary winding L_S of transformer are depending up of turns ratio determine from (2):

$$L_{Sj} = \frac{L_P}{p_j^2}, j = \sum_{i=1}^{9} i \qquad (2)$$

TABLE II
THE VALUES OF INDUCTANCE OF SECONDARY WINDING L_S
OF TRANSFORMER FOR DIFFERENT VALUSES OF TURNS RATIO P

j	L_{Sj} [H]
1	1,23
2	0,78
3	0,54
4	0,4
5	0,3
6	0,24
7	0,19
8	0,16
9	0,13

V. SIMULATION RESULTS

The simulation model of FB converter on Fig.4 was created in programme OrCAD Capture CSI to verify the theoretical assumes. Parameters:

Switching frequency f_S = 5 kHz,
Input voltage U_{PV} = 8 – 12 V,
Output voltage U_{bat} = 6V,
Inductance of primary side of TR L_P = 100mH,
Inductance of secondary side of TR L_S = 130 – 300mH.
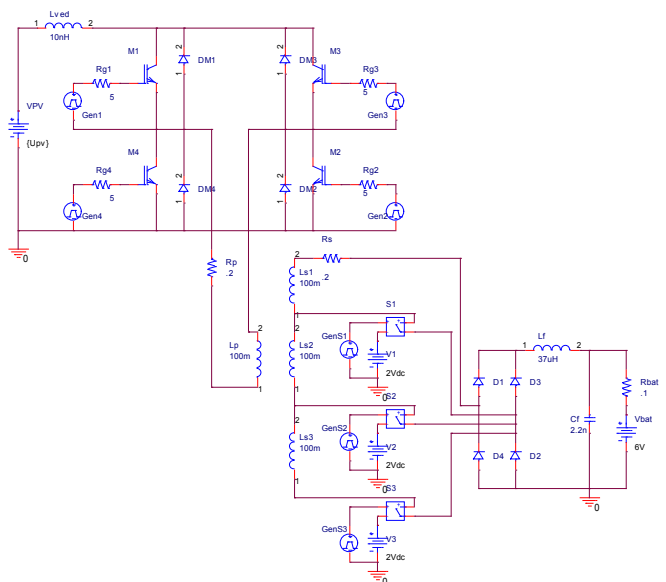
Fig.4. Simulation model of FB converter.

For illustrative purpose the simulation was created only for three values of input voltage namely U_{PV} = 12V, 10V and 8V and corresponding inductance of secondary side of transformer L_S, see Tab.1, Tab.2 respectively. For simulation

proposes the diagonally opposite switches M1, M2, or M3, M4 are turned on and off simultaneously.

Fig.6. shows the waveforms of input voltage U_{PV}, load current I_Z, voltage of the primary U_P and secondary U_S side of transformer. It can be seen that for changes in irradiance or more precisely input voltage U_{PV} the voltage on the secondary side of transformer U_S stay constant. This constant value of secondary voltage U_S is ensured by change of turns ratio. The control system senses the change of input voltage U_{PV} and set up the turns ratio *p* by switching the proper transformer tap.

a)

b)

c)
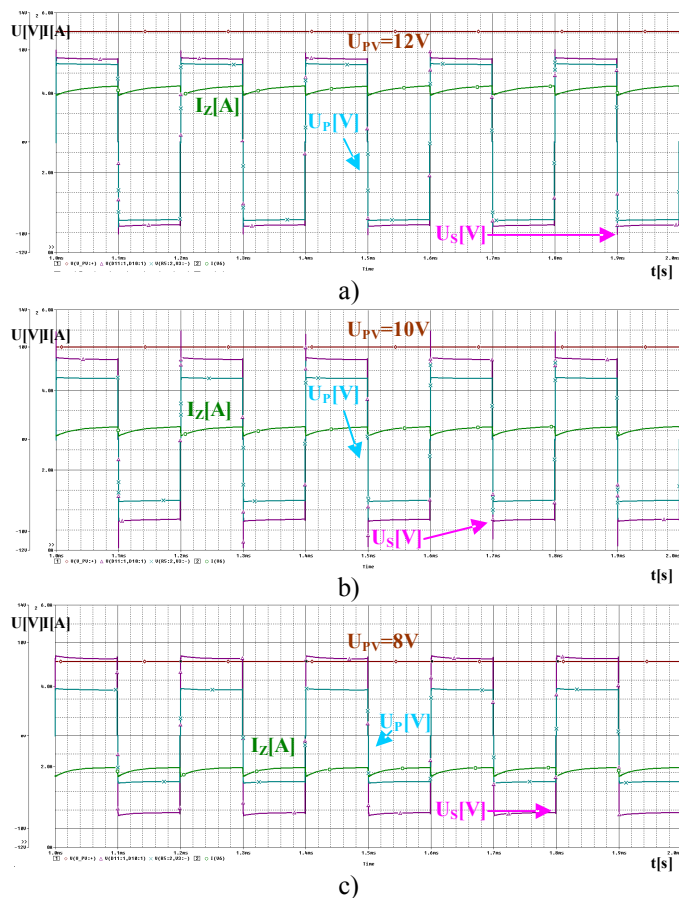
Fig.5. The waveforms of input voltage U_{PV}, load current I_Z, voltage of the primary U_P and secondary U_S side of transformer at different levels of irradiance or more precisely input voltage U_{PV} a) U_{PV} = 12V b) U_{PV} = 10V c) U_{PV} = 8V.

From Fig.5. it can be seen that the value of secondary voltage U_S is approximately 8V instead of 14V. This difference is because of battery voltage on the output. The battery voltage U_{bat} acts as counter-voltage against the secondary voltage of transformer U_S.

VI. CONCLUSION

The most popular material use to make the solar cell is silicium (Si). The Si is also major factor in long-term energy recovery and high cost of PV cells. One way to reduce the long –term energy recovery and so high cost of PV cells is proposed PV system. This new concept of PV system ensured of utilize the full range of energy supplied from the solar cell

by proper switching particular tapped winding on the secondary side depending up the irradiance or more precisely input voltage U_{PV}.

REFERENCES

[1] Fundamental of photovoltaic materials. In: National Solar Power Research Institute, Inc. 12/21/98, pp. 10.

[2] Kováč, D., Kováčová, I.: EMC Aspect as Important Parameter of New Technologies. In: New Trends in Technologies: Control, Management, Computational Intelligence and Network Systems, Publisher: Sciyo, November 2010, pp. 305-334, ISBN 978-953-307-213-5.

[3] Kováč, D., Kováčová, I.: Power transistors MOSFET and IGBT (in Slovak), Košice: Elfa, 1996, 117 pages.

[4] Kováčová, I., Kováč. D.: Safeguard circuits of power semiconductor parts. In: Acta Electrotechnica et Informatica. Vol. 3, No. 3 (2003), pp. 44-51. ISSN 1335-8243.

[5] Kováč, D., Kováčová, I.: Modelling and measuring of the electronic circuits. Košice: Elfa, 1996, 92 pages.

[6] Kováčová, I., Kováč. D., Oetter, J.: Applied electronics: recommendations for exercises I (in Slovak). Košice: Akris, 2001, 94 pages, ISBN 80-968666-0-5.

[7] Kováčová, I., Kováč, D.: Industrial Electrical Engineering – Modeling of converters, 1st part. FEI TU Košice, 2011, 66 pages, ISBN 978-80-553-0617-9

[8] Kováčová, I., Kováč, D.: Industrial Electrical Engineering – Modeling of converters, 2nd part. FEI TU Košice, 2011, 68 pages, ISBN 978-80-553-0618-6

[9] Kováč, D., Vince, T., Molnár, J., Kováčová, I.: Modern Internet Based Production Technology. In: New Trends in Technologies: Devices, Computer, Communication and Industrial Systems, Book edited by: Meng Joo Er, Publisher: Sciyo, November 2010, pp. 145 – 164, ISBN: 978-953-307-212-8

A single chip for UWB applications consisting of two LNAs and differential amplifier

Martin LIPTAJ, Branislav HRUŠOVSKÝ, Radovan BLICHA

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

martin.liptaj@tuke.sk, branislav.hrusovsky@tuke.sk, radovan.blicha@tuke.sk

Abstract—We demonstrate the design and implementation of the two Low noise amplifiers and differential amplifier on the single chip manufactured in the 0.35 μm SiGe BiCMOS technology. This chip will be a part of the set of ASIC circuits designed for implementation in the recent UWB radar system architectures.

Keywords—ASIC, Differential Amplifier, Low Noise Amplifier, SiGe BiCMOS, Ultra-WideBand (UWB)

I. INTRODUCTION

Noise Factor and Noise Figure are basic parameters, which describe or characterize each system in terms of noise. For most applications the small Noise Factor is important assumption. Typical application for this assumption is Low noise amplifier (LNA), which require a high gain and a flat frequency response over the entire UWB frequency band as well. Additional priorities for the LNA are impedance matching and linear phase response. Another challenge is the necessary dynamic range of the LNA due to the presence of in-band interferers along with a moderate power consumption. Because of these facts, the LNA amplifiers as input stages of miscellaneous devices (mobile phones, radar systems and many more) are used [1]. If noise distortion is small (i. e. the value of Noise Factor is also small), the results of post-processing are then more effective [2]. The topology inspirations for these LNA amplifiers were similar to project described in [3].

The differential amplifier is one of the most versatile circuits used in analog circuit as well as digital design. These are widely used in electronics industry and are generally preferred over their single-ended counterparts because of their better common-mode noise rejection, reduced harmonic distortion, and increased output voltage swing. Differential amplifiers are used to amplify analog as well as digital signals, and can be used in various implementations to provide an output from the amplifier in response to differential inputs. It is very compatible with integrated circuit technology and serves as the input stage to most of applications [4]. They can be comfortably adapted to function as an operational amplifier, a comparator, a sense amplifier/tripler [5] as well as a front-end buffer stage for another circuits [6] as for example for divider or as a BalUn (BALanced-UNbalanced), i.e. device that convert electrical signals that are balanced about ground (differential) to signals that are unbalanced (single-ended) and vice versa.

Chip consist of two identical LNA amplifiers and one differential amplifier. It is designed in 0.35 μm SiGe BiCMOS technology from Austriamicrosystems (AMS), Austria[7].

II. LOW NOISE AMPLIFIERS

Low noise amplifiers operate in Class A, which is characterized by a bias point more or less at the center of maximum current and voltage capability of the used device, and by RF current and voltages, that are sufficiently small relative to the bias point that the bias point does not shift. These LNA amplifiers are a highly compact amplifiers based on a cascode topology with resistive feedback in combination with an emitter follower stage (figure 1).

This circuit was designed in order to operate with maximal frequency bandwidth, with optimal gain and minimal noise figure. The main noise contributor to the overall noise figure is input transistor Q_2 which is connected with common emitter. The output buffer is an emitter follower stage which provides broadband 50 Ω match to the output. This buffer consists of transistor Q_4 and resistor R_6. Broadband operation in terms of
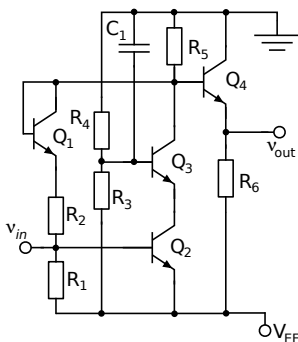
Fig. 1.   Schematic of the low noise amplifier
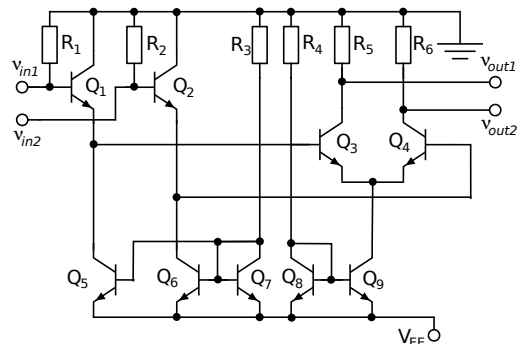
Fig. 2.   Schematic of the differential amplifier

noise performance, flat gain response, bandwidth and power matching has been achieved with a simple resistive shunt feedback in combination with a diode DC level shifter which improves the large signal behavior of the design. The DC level shifter in the shunt feedback uses the base emitter diode of transistor Q_1 which is biased via resistors R_1 and R_2. Transistor Q_3, which is connected as common base to matching between input stage (transistor Q_2) and output buffer stage (transistor Q_4) is used. Base of this transistor is connected to virtual ground created using resistors R_3 and R_4 and MIM capacitor C_1.

The result of the post-layout AC analysis without output load connection of this LNA is shows in Fig. 3. As can be seen, the simulated -3dB bandwidth is around 5,9 GHz due to parasitic parameters. In ideal case the simulated -3dB bandwidth is more than 8 GHz. Simulated gain is more than 20 dB in the frequency range of interest and Noise figure is less than 3 dB.
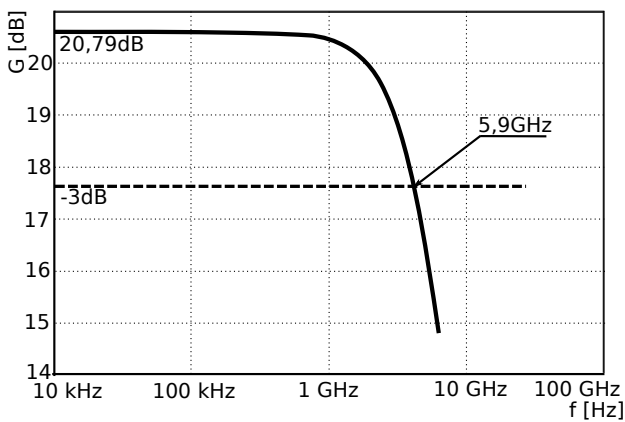
Fig. 3.   AC analysis of the low noise amplifier designed in 0,35 μm SiGe BiCMOS technology
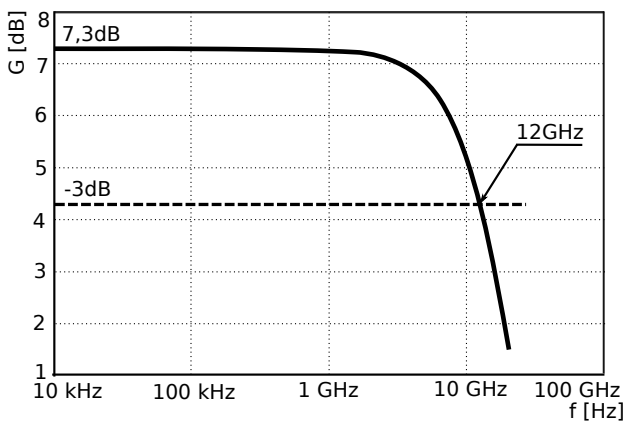
Fig. 4.   AC analysis of the differential amplifier designed in 0,35 μm SiGe BiCMOS technology

III. DIFFERENTIAL AMPLIFIER

The proposed differential amplifier is shown in Fig. 2. This circuit was designed in order to operate with maximal frequency bandwidth, with optimal differential gain and maximal common mode suppression. Pair of transistors Q_3 and Q_4 forms core of differential amplifier. The transistors Q_1 and Q_2 with the resistors R_1 and R_2 form the input stage of differential amplifier. In the emitter circuit of the differential

amplifier, a Widlar current source is used. The current of this source formed by Q_8, Q_9 and R_4 create bias current for the differential stage. The biasing of the transistors Q_3 and Q_4 is solved by current mirrors consisting with the transistors Q_5, Q_6, Q_7 and R_3 [8].

The result of the AC analysis of this differential amplifier is shows in Fig. 4. As can be seen, with connected 50 Ω load on the output, the simulated -3dB bandwidth is around 12 GHz. Simulated gain is more than 7 dB in the frequency range of interest and CMRR was foun out better than 50 dB.
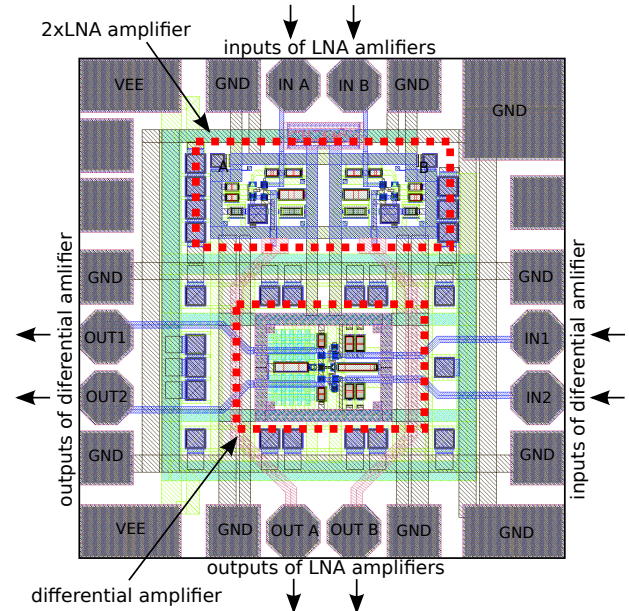
Fig. 5.   Layout of the differential amplifier and pair of low noise amplifier in 0,35 μm SiGe BiCMOS technology
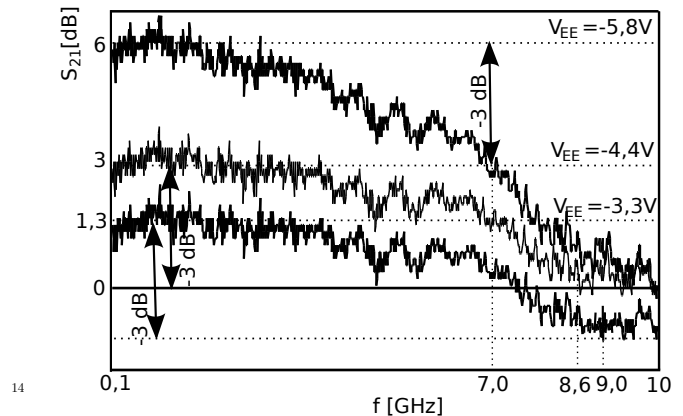
Fig. 6.   Measured gain S_{21} of the differential amplifier for supply voltages V_{EE} -3,3 V, -4,4 V and -5,8 V

IV. IMPLEMENTATION AND MEASUREMENT RESULTS

The layout of the final chip is shows in Fig. 5. The chip consists of the differential amplifier and also from the low noise amplifier pair. This concept was used because of maximal chip area utilization. The selected chip dimension for implementation of ASIC was 850x830μm. Die area of LNA is 100 x 180 μm and die area of differential amplifier is equal to 180 x 230 μm. In the final manufactured wafer (Fig. 7) the amplifiers (LNA and differential) in 0,35 μm SiGe BiCMOS technology from AMS were designed.

The on-chip preliminary measurements have been performed by a probe station and a Network Analyzer Rohde&Schwarz R&S®ZVA40, in 50 Ω environment. The measurements only on differential amplifier were realized so far. In Fig. 6 the gain (S_{21}) of the proposed differential amplifier is shown. Both amplifiers (LNA and differential) were designed for the supply voltage -3,3V with the common current consumption 29 mA (LNA amplifiers and differential amplifier have the current consumption less than 14 mA and 15 mA, respectively). Under these conditions, measured -3 dB cut-off frequency is around 9 GHz and the whole chip consumes 28 mA of current. The single ended measured gain is around 1,3 dB in the frequency range of interest. It should be noted that the measurement of the differential amplifier has been done in single-ended mode and that the structure is designed for differential operation. Therefore the single-ended gain is 6 dB lower than the equivalent output values of a true-differential signal. Consequently for differential output the total gain will be 7,3 dB as it was found out from simulation (Fig. 4). The broadband input match of differential amplifier is better than 15 dB, output return loss lies below 10 dB and reverse voltage gain is better than 50 dB in the frequency range of interest.

In order to test the limits of the technology and robustness of the design several measurements of the amplifier gain with different supply voltages was performed (see Fig. 6). Under these conditions (V_{EE}=-5,8 V), output single-ended gain is around 6 dB (eqvivalent to 12 dB in differential mode) and -3 dB bandwidth was decreased to 7 GHz. The current consumption of the whole chip is 78 mA.

In the future work it will be important to cancel or suppress the parasitic components of the amplifier building blocks wirings through the sophisticated selection of the particular interconnection structures. This means for example that width of metals must be thinner, overlap of metals must be minimal etc. On the other hand the metal connections must be in compliance with the design rules, therefore for example the metal connection must be of the sufficient width and with the sufficient space between them. Consequently the important for the design will be to find the trade-off among these contradictions. Our future work will also focus on low noise amplifiers measurements.

V. CONCLUSIONS

A low noise amplifiers and a bipolar differential amplifier for UWB applications using 0,35 μm SiGe BiCMOS technol-

ogy from AMS are presented. The results from the simulation and from the measurements of the differential amplifier under test were similar. The difference was only in the achieved frequency bandwidth, where measured cut-off frequency was 9 GHz i.e. smaller as simulated 12 GHz one. This is most likely caused by parasitic components between metal connections on the chip, which will in future be important to cancel or suppress. Measurement of LNA amplifiers will be performed in near future.

REFERENCES

[1] M. Liptaj, "Návrh ASIC obvodov pre širokopásmové aplikácie," *Minimum thesis*, 2010, Technical University of Košice.
[2] B. Hrušovský, "Návrh a implementácia algoritmov pre maskovanie chýb prenášaných snímok v bezdrôtovom prostredí," *Minimum thesis*, 2010, Technical University of Košice.
[3] J. Dederer, S. CHhartier, T. Feger, U. Spitzberg, A. Trasser, and H. Schumacher, "Highly Compact 3.1 - 10.6 GHz UWB LNA in SiGe HBT technology," *Microwave Integrated Circuit Conference*, pp. 327 – 330, 2007, EuMIC 2007, Munich.
[4] S. Chartier, B. Schleicher, F. Korndorfer, S. Glisic, G. Fischer, and H. Schumacher, "A fully integrated fully differential low-noise amplifier for short range automotive radar using a SiGe:C BiCMOS Technology," *Microwave Integrated Circuit Conference*, pp. 407 – 410, 2007, EuMIC 2007, Munich.
[5] M. Liptaj, M. Kmec, P. Galajda, and J. Sachs, "Recent SiGe frequency tripler development for a new UWB system architectures," *Research in Telecommunication Technologies 2010 : the 12th International Conference*, September 2010, Velké Losiny, Czech Republic. Ostrava : VŠB-TU, ISBN 978-80-248-2261-7.
[6] F. Centurelli, R. Luzzi, P. Marietti, G. Scotti, P. Tommasino, and A. Trifiletti, "An active balun for high-CMRR IC design," *Gallium Arsenide and Other Semiconductor Application Symposium*, pp. 621 – 624, Oct. 2005, EGAAS 2005,.
[7] www.austriamicrosystems.com, "website of company," *available online:*, http://www.austriamicrosystems.com.
[8] M. Liptaj, P. Galajda, and M. Kmec, "Recent Fully Differential Amplifier in 0,35 μm SiGe BiCMOS technology for UWB applications," 2011, Unpublished.
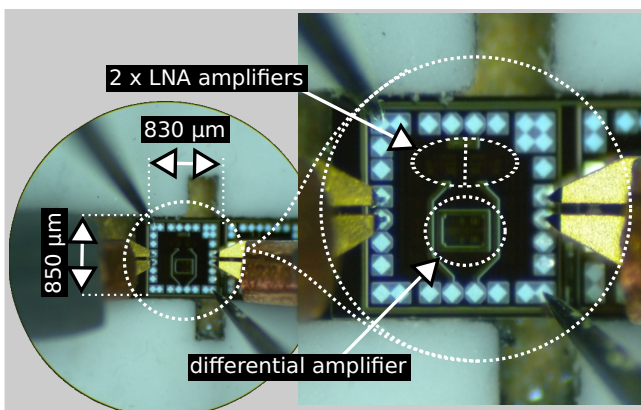
Fig. 7. Photo of the measured chip wafer in 0,35 μm SiGe BiCMOS technology

An Analysis of OFDM, Precoded OFDM, OFDMA and Single Carrier – FDMA in Cellular Systems

[1]*Radovan BLICHA,* [2]*Branislav HRUŠOVSKÝ,* [3]*Martin LIPTAJ*

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]radovan.blicha@tuke.sk, [2]branislav.hrusovsky@tuke.sk, [3]martin.liptaj@tuke.sk

Abstract—The purpose of the present paper is to compare OFDM (Orthogonal frequency division multiplexing), precoded OFDM, OFDMA (Orthogonal Frequency Division Multiple Access) and SC-FDMA (Single Carrier – frequency division multiple access) modulation schemes for wireless communication systems. In this paper we present a new analysis of broadband cellular systems. These techniques are based on precoding the constellation symbols, which use the spreading matrix some sequences. For simulation was employed Walsh-Hadamard sequences and compare with "classic" OFDM and SC-FDMA. The proposed codes have good auto- and cross- correlation function.

Keywords—LTE, SC-FDMA, OFDM, OFDMA, Spread sequences, Walsh-Hadamard matrix, PAPR.

I. Introduction

LTE (Long Term Evolution) is a next generation mobile system from the 3GPP with a focus on wireless broadband. LTE is based on Orthogonal Frequency Division Multiplexing (OFDM) with cyclic prefix (CP) in the downlink, and on Single-Carrier Frequency Division Multiple Access (SC-FDMA) with cyclic prefix in the uplink [1].

SC-FDMA is a new technology for high data rate uplink communication and has been adopted by 3GPP for it next generation cellular system, called LTE. LTE implements OFDM for its downlink and SC-FDMA for its uplink. SC-FDMA is a modified form of OFDM with similar performance and complexity. The primary advantage of OFDM over single-carrier schemes is its robustness against multipath signal propagation, which makes it suitable for broadband systems. On the other hand, SC-FDMA brings additional benefit of low peak-to-average power ratio (PAPR) compared to OFDM.

In this paper, we compare the PAPR of transmitter systems OFDM, precoding OFDM and SC-FDMA. This paper provides an analysis of these new techniques for the reduction of PAPR.

The paper is organized as follows: Section 2 gives an introduction to the topic of OFDM and SC-FDMA. Section 3 is about the Spreading sequences. Section 4 presents computer simulations of used system models, complexity and implementation issues. Finally, the conclusions are made in last section of this paper.

II. Why OFDM and SC-FDMA

OFDM is very popular modulation scheme for wideband digital communication. It has been included in digital audio/video broadcasting (DAB/DVB) standards in Europe, successfully applied to high-speed digital subscriber line (DSL) modems in the US, and recently proposed for digital cable television systems and local area mobile wireless networks such as the IEEE802.11a and the HiperLAN [2].

OFDM is a most favored technique for broadband wireless system due to susceptibility to signal dispersion under multipath conditions. OFDM can also be viewed as a multi-carrier narrowband system where the whole system bandwidth is split into multiple smaller subcarriers with simultaneous transmission. Simultaneous data transmission and reception over these subcarriers are handled almost independently. Each subcarrier is usually narrow enough that multipath channel response is flat over the individual subcarrier frequency range, i.e. frequency non-selective. Another way to look at is that an OFDM symbol time is much larger than the typical channel dispersion. (Even though OFDM symbol duration is much larger than channel dispersion). Hence OFDM is inherently susceptible to channel dispersion due to multipath propagation [1].

OFDM is a method of digital modulation. The system model for this access method can be seen in Figure 1. The input binary data (signal) is split into several narrowband channels at different frequencies. Standard OFDM transmitter generates N-QAM or QPSK modulated subcarriers that are orthogonal. OFDM use IFFT at the transmitter and FFT at the receiver. To avoid inter-block interference (IBI) between transmitter's symbols is addition cyclic prefix (CP).

OFDMA is a multi-user OFDM that allows multiple access on the same channel (a channel being a group of evenly spaced subcarriers, as discussed above). WiMAX uses OFDMA, extended OFDM, to accommodate many users in the same channel at the same time.

OFDMA distributes subcarriers among users so all users can transmit and receive at the same time within a single channel on what are called subchannels. What's more, subcarrier-group subchannels can be matched to each user to provide the best performance, meaning the least problems with fading and interference based on the location and propagation characteristics of each user.
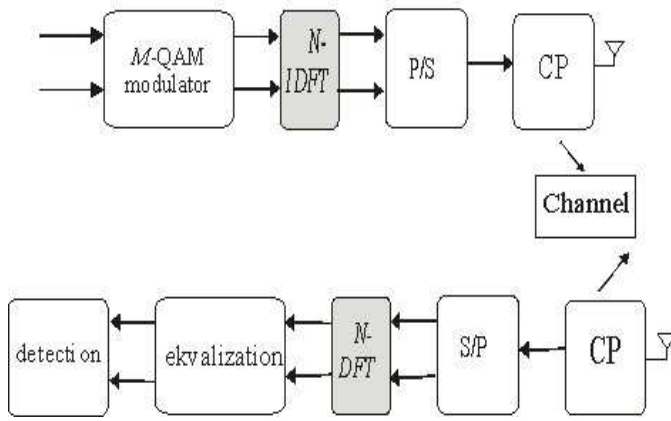
Fig. 1. Block Diagram of the OFDM Systems.

Another system, which we will describe, is a SC-FMDA. This is a new uplink air interface technique for 4G wireless multiple access technology, which utilizes single carrier modulation, DFT-spread orthogonal frequency multiplexing, and frequency domain equalization. It has a similar complexity, structure and performance as OFDM.

With respect to resource allocation, we consider that a subset of subcarriers comprises a chunk, and one or more chunks are allocated to each user, since coordinating individual subcarriers is too burdensome. Thus, the number of subcarriers in a chunk is regarded as a minimum resource unit for subcarrier allocation. As mentioned earlier, there are two types of subcarrier mapping: Localized FDMA (L-FDMA) and Interleaved FDMA (I-FDMA). A chunk in L-FDMA consists of consecutive subcarriers. Subcarriers in a chunk of IFDMA are distributed over the entire bandwidth with equal frequency spacing [5].

One prominent advantage over OFDM is that the SC-FDMA signal has lower PAPR because of its inherent single carrier structure [3]. SC-FDMA has drawn great attention as an attractive alternative to OFDM, especially in the uplink communications where lower PAPR greatly benefits the mobile terminal in terms of transmit power efficiency. It is currently a strong candidate for uplink multiple access scheme in LTE of 3GPP [4].

III. SEQUENCES AND PRECODING

One of the most well known precoding matrices is application differences codes e.g., Walsh-Hadamard (WH). The Walsh–Hadamard transform (WHT) and discrete Fourier transform (DFT) are highly practical value for representing signals, images and mobile communications for orthogonal code designs.

Precoded OFDM consists of using a precoding matrix P that spreads the energy of symbols over the subcarriers allocated to the user. In this letter the Hadamard Matrix is defined recursively as bellow:

$$c_n = \begin{bmatrix} c_{(n-1)} & c_{(n-1)} \\ c_{(n-1)} & -c_{(n-1)} \end{bmatrix}$$

The Walsh-Hadamard code is an error correcting code over a binary alphabet that allows reconstruction of any codeword if less than half its bits are corrupted. Furthermore, being a locally decodable code, the Walsh-Hadamard code is an important tool in computational complexity theory, and can be particularly used in the design of probabilistically checkable proofs. The Walsh-Hadamard code, calculated by the Walsh function, is used to uniquely define individual communication channels. Walsh codes are mathematically orthogonal codes. As such, if two Walsh codes are correlated, the result is intelligible only if these two codes are the same. As a result, a Walsh-encoded signal appears as random noise to a SC-FDMA capable mobile terminal, unless that terminal uses the same code as the one used to encode the incoming signal [6].
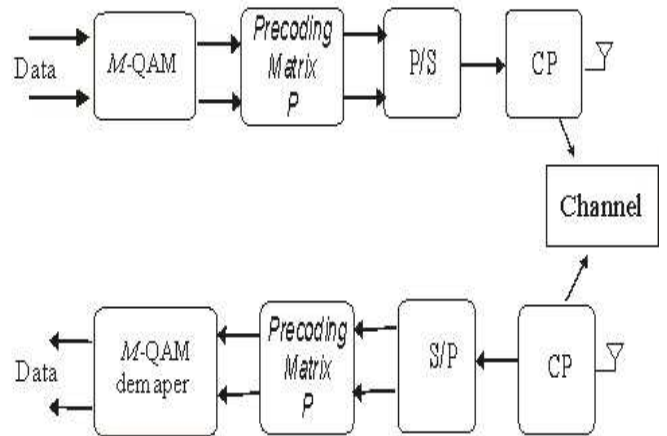
Fig. 2. Block Diagram of the transmitter systems with precoding matrix.

In this paper we compare PAPR for difference transmitters systems. The PAPR of signal is given by:

$$PAPR = \frac{max|x_n|^2}{E\{|x_n|^2\}}$$

(1)

where E{.} denotes taking the expected value and $E\{|x_n|^2\}$ is equal to the variance σ_x^2, since the symbol are zero-mean. Statistics of the PAPR of an OFDM signal can be given in terms of its clipping probability or its complementary cumulative distribution function (CCDF). The CCDF, for an OFDM signal is defined as:

$$P(PAPR > PAPR0) = 1 - (1 - e^{-PAPR0})^N$$

(2)

where PAPR0 is the clipping level. This equation can be interpreted as the probability that the PAPR of a symbol block exceeds some clip level PAPR0 [7].

IV. SIMULATION RESULTS

Extensive simulations have been performed in order to evaluate the performance of the proposed system as compared with the conventional OFDM system. Simulation results of WHT-precoded system are also included for comparison.

In the present work, we have considered QAM modulation based on OFDM system. Figure 3 shows the comparison of CCDF for the OFDM, precoded OFDM, OFDMA and SC-

FDMA transmitter systems, for N=64, 256. All simulation results have been randomly generated and modulated into 16-QAM. Soft Limit, Saleh of HPA and IBO=2, 8 were another set parameters used for simulations.

CCDF (N=64)

Fig. 3. Transmitter systems OFDM, Precoded OFDM, OFDMA, SC-FDMA, (N=64)

CCDF (N=64)

Fig. 4. Transmitter systems OFDM, Precoded OFDM, OFDMA, SC-FDMA, (N=64), HPA Saleh

From Figure 3, with N=64, we can observe already mentioned compared transmitter systems. From this simulation we can notice that the precoded OFDM outperforms the OFDM system.

From Figure 4, with N=64, we can observe already mentioned compared transmitter systems. From this simulation we can notice that the precoded OFDM outperforms the SC-FDMA system.

In Figure 5, with N=256, at the clip rate of 10-3, the precoded OFDM gives a performance gain of less than 1dB while SC-FDMA schema can achieve PAPR reduction.

In Figure 6, with N=256, at the clip rate of 10-3, the precoded OFDM gives a performance gain more than 1dB

while SC-FDMA schema can achieve PAPR reduction. As we can see from simulation the precoded transmitter system brings additional benefit of low PAPR compared to another transmitters system.

CCDF (N=256)

Fig. 5. Transmitter systems OFDM, Precoded OFDM, OFDMA, SC-FDMA, (N=256)

CCDF (N=256)

Fig. 6. Transmitter systems OFDM, Precoded OFDM, OFDMA, SC-FDMA, (N=256), HPA Saleh

As the number of subcarriers, N increases, the PAPR increases (the CCDF curves shift to the right). PAPR reduction achieved by using sequence codes.

One of possibilities in the future is using Spread sequences and nonlinear amplifier.

V. CONCLUSION

In this paper, we have reviewed the fundamentals of three techniques which are potential candidates for the uplink and downlink of future cellular communications systems: OFDM, Precoded OFDM, OFDMA and SC-FDMA.

The primary advantage of OFDM over single-carrier schemes is its robustness against multipath signal propagation. On the other hand, thanks to its single-carrier signal structure,

SC-FDMA brings additional benefit of low PAPR compared to OFDM system. In the future, using various sequences codes brings outperforms over conventional transmitter systems.

REFERENCES

[1] Ixia Worldwi de Headquarters, "SC-FDMA Single Carrier FDMA in LTE", November 2009, 915-2725-01.

[2] ETSI, "Broadband Radio Access Networks (BRAN), HIPER-LAN Type 2 technical specification Part 1 - physical layer", October 1999, DTS/BRAN030003-1.

[3] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar and B. Eidson, "Frequency Domain Equalization for Single-Carrier Broadband Wireless Systems", *IEEE Commun. Mag.*, April 2002, vol. 40, no. 4, pp. 58–66.

[4] 3rd Generation Partnership Project (3GPP), Technical Specification Group Radio Access Network, Physical Layer Aspects for Evolved UTRA, http://www.3gpp.org/ftp/Specs/html-info/25814.htm.

[5] J. Lim, H. G. Myung, K. Oh, D. J. Goodman, "Channel-Dependent Scheduling of Uplink Single Carrier FDMA Systems", *IEEE Transactions on Microwave Theory and Techniques,* 2006, 1-4244-0063-5/06.

[6] H. Sari, C. Ciochina, D. Motier, "An Analysis Of OFDMA, Precoded OFDMA And SC-FDMA For The Uplink In Cellular Systems", European Transactions on Telecomunications, august 2008, vol. 19, Issues 5, p. 581-588.

[7] Y. K. Min, V. Jeoti, "A Novel Signail Independent Technique for PAPR Reduction in OFDM Systems", *IEEE – ICSCN, MIT Campus, Anna University, Chennai, India*, 22-24 February 2007, pp.308-311.

14

Analysis of Servodrive Parasite Nonlinearities

^{1}*Matúš HRIC,* ^{2}*Karol KYSLAN,* ^{3}*Peter NGUYEN*

^{123}Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

^{1}matus.hric@tuke.sk, ^{2}karol.kyslan@tuke.sk, ^{3}peter.nguyen@tuke.sk

Abstract— **The paper describes basic terms occurring in area of servodrive precise control like a backlash and friction. The knowledge is necessary for further analysis of parasite nonlinearities by identification methods that present key issues in control structures of servodrive precise control .**

Keywords—**Backlash, Friction, Identification.**

I. INTRODUCTION

Increasing requirements to position control precision or of motion of material in a production process impose conditions to modernization of technologies, like using more precise optical position sensors, higher quality drives, more precise gears and more powerful control processors that react to all motion aspects. The compact servodrives merging in one unit the machine, power converter with control and gear present development trend over the last years. Due to nonlinearity of the servodrive, its part, and very often also of the driven mechanism applications, the nonlinear structures compensating unwanted influences inputting into motion control, present the basic drive structures. Identification and following compensation became a standard for precise servodrives control. Currently, the issues of cost cutting by integration and eventually substitution of special sensors that increase expenditures and final price are topical. In this case identification and following compensation of system disturbances in control structure can be performed without its sensing.

II. SERVODRIVE (ACTUATOR)

The term "servodrive" is explained by various definitions but from view of industrial utilization of servodrives the next one seems to be the best:

Servodrive as industrial equipment fulfill the task of the initiate the driven mechanism (machine) by a prescribed way into specified motion so that the required technological operation or technological process would be accomplished [1].
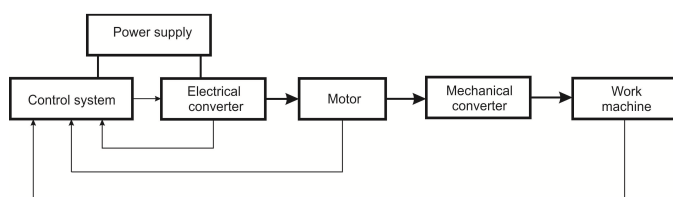
Fig 1. Servodrive structure

The servodrive can be looked on as a relatively autonomous system with internally organized combination of subsystems. It is created by individual subsystem properties where their number and type depends on the particular type of servodrive. Its typical arrangement is shown in Fig.1 where the information flow is marked by a thin line and power (energy) flow is marked by a thick line.

III. PARASITE NONLINEARITIES

Parasite nonlinearities influence negatively the system properties that result in deterioration of control circuit properties. Their identification and mathematical description presents a basis for following compensation. The friction model according to ref. [2] and based on [2] consists of four components: static friction, viscose friction, coulomb friction and so called Stribeck effect.
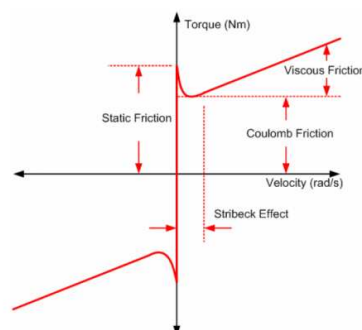
Fig 2. Friction model

1) <u>Static friction</u> occurs in case when objects are at rest. If the external torque is larger than static friction torque, the object starts to rotate. This static friction can be described as follows:

$$M = \begin{cases} M_e & ak\ \omega = 0\ a\ |M_e| < M_s \\ M_s sgn(M_e) & ak\ \omega = 0\ a\ |M_e| \geq M_s \end{cases} \quad (1)$$

2) <u>Coulomb friction</u> (called also dry friction) is presented by torque acting against the object motion, at which it actuates the object by a constant torque. The torque M_c corresponding to the Coulomb friction is described by:

$$M_c = \mu M_N \quad (2)$$

where μ is Coulomb friction coefficient, M_N is normal torque.

3) <u>Viscous friction</u> is presented by a linear function of the torque proportional to the speed. Dependence of the viscous friction torque on the angular is described by the equation:

$$M_v = k_v \omega \qquad (3)$$

where k_v represents viscous friction coefficient and ω the angular velocity.

4) <u>Stribeck effect</u> follows up from using liquid or solid lubricant across contact surfaces of mutually moving mechanical parts in the drive. This causes a drop of the friction at larger speeds (Fig. 1) up to so called Stribeck speed that depends on used lubricant.

Friction model of M_t described in ref. [2] is approximated to the Armstrong friction model described as:

$$M_t(\omega) = k_v\omega + M_c sgn(\omega) + [M_s - M_c]sgn(\omega)e^{-\left(\frac{\omega}{v_s}\right)^2} \quad (4)$$

where k_v is friction coefficient of the viscous friction, M_c is Coulomb friction torque, M_s is maximal value of static friction torque , v_s is Stribeck speed and ω is angular speed.

Armstrong model contains all important friction torque components like static, Coulomb, and viscous frictions and Stribeck effect and it represents the best the wide range of friction nonlinearities that are available at control of servodrives.

IV. BACKLASH

The backlash is defined as a gap among adjacent mechanical part of gear grades in the gear or as a gap which the area for the tooth exceeds the tooth width at the pitch circle, [3]. Backlash influenced negatively precise control of servodrives that contain a gear mechanism. The characteristic of static backlash and object of demonstration is shown in Fig. 3.
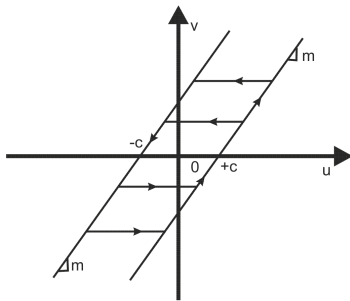
Fig. 3. Characteristic of mechanical hysteresis

The backlash in the graph is represented by two parallel lines – one is increasing and the second one decreasing that are connected by horizontal segments the width of which presents the very backlash. The increasing line is active if the input $v(t)$ and output $u(t)$ increase, i.e.: $u(t) = mv\big((t) + c\big)$, where $m > 0$, $\dot{v}(t) > 0$.

The decreasing line is active if $v(t)$ and simultaneously $u(t)$ decrease, i.e.: $u(t) = mv((t) - c_p)$, $\dot{v} < 0$.

If the tooth of gear-wheel is inside the gap between the teeth of engaged gear-wheel, then $\dot{u}(t) = 0$.

The mathematical representation of backlash is as follows:

$$\dot{u}(t) = \begin{cases} mv(t) & \text{if} \quad \dot{v}(t) > 0 \text{ and } u(t) = m(v(t+c), \\ & \text{or} \quad \dot{v}(t) < 0 \text{ and } u(t) = m(v(t-c) \quad (5) \\ 0 & \text{in other cases} \end{cases}$$

The response of backlash to sinus signal is shown in Fig. 4.
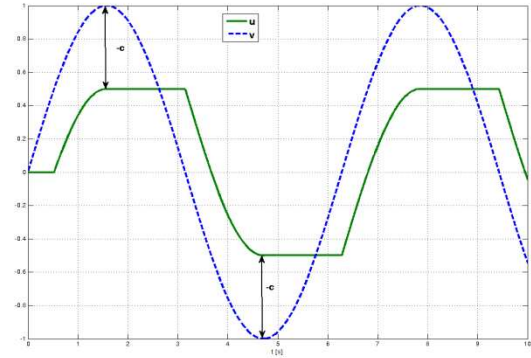
Fig. 4. Response of output shaft

If dynamic effect cannot be neglected, one has to use so called Sandwiched Backlash model, [5]. This model takes into consideration motor, load, and torque here the transfer from the motor to the load is modeled by a damper and a spring according to the ref. [6] as follows: let 2α is Backlash value, k is elasticity constant, b is a damping constant, M is torque, $\theta_1(t)$ – angle of motor rotor displacement (gear input shaft), $\theta_3(t)$ – angle of je gear output shaft with the backlash, $\theta_2(t)$ – angle of driven machine (load).

Let's define:
- total angular displacement $\theta_d(t) = \theta_1(t) - \theta_2(t)$,
- shaft twisting $\theta_s(t) = \theta_1(t) - \theta_3(t)$
- and angle of backlash $\theta_b(t) = \theta_3(t) - \theta_2(t)$.

The shaft torque M_h is:

$$M_h(t) = k\theta_s + b\dot{\theta}_s = k(\theta_d - \theta_b) + b(\dot{\theta}_d - \dot{\theta}_b)$$

and position of backlash is conforms the following dynamics:

$$\dot{\theta}_b = \begin{cases} max\left(0, \dot{\theta}_d + \frac{k}{b}(\theta_d - \theta_b)\right), & \theta_b = -\alpha \\ \dot{\theta}_d + \frac{k}{c}(\theta_d - \theta_b) & |\theta_b| < \alpha \quad (6) \\ min\left(0, \dot{\theta}_d + \frac{k}{b}(\theta_d - \theta_b)\right), & \theta_b = \alpha \end{cases}$$
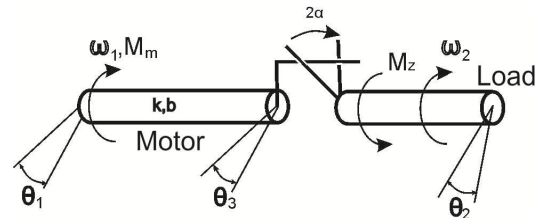
Fig. 5. Sanwiched Backlash model

V. MECHANICAL COMPENSATION OF BACKLASH

Due to high requirements to precise positioning mechanisms of robots, manipulators, or NC machines, the build up precise gears (reductors) were developed there having a very small value of mechanical hysteresis. The gears of the Spinea Twinspin type Fig.6 integrate in them a function of reductor with function of bearing, [7].
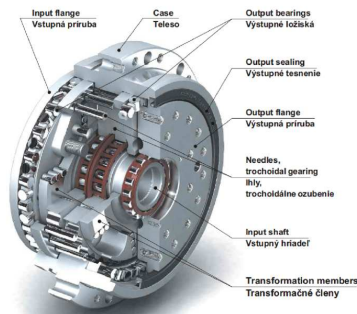
Fig. 6. Spinea Twinspin type gear

The components of the gear are installed in a prestrained state due to linearization of torque characteristics and minimalization of its backlash.

VI. IDENTIFICATION

The backlash identification that is necessary for unwanted effect compensation can be realized by several ways. In ref. [8] there is shown an identification based on two methods, namely:

1) *Indirect identification* of backlash measures a dynamic system response to very small torque impulse generated by the drive in the opposite rotating direction. If the torque impulse is smaller than the gap among engaged gear teeth, the output shaft shows any reaction. The magnitude of the torque impulse is increasing up to the time when the tooth starts to gear in the opposite direction.

2) The second method (called also a *quasi-stationary method*) is applied at small speeds where it is satisfactory to neglect moment of inertia and it is sufficient to consider only the friction nonlinearity. The method of backlash identification without any position sensor on the output is based on following a torque component of the current i_{sq} (the author in the reference uses the AC motor). At position change (Fig. 7), where θ_{in} is motor position angle, θ_{out} is output gear shaft angle there is shown a course of the gear output shaft speed. In time t_1 there is a change of position, the current traverse from negative values to positive values. The period between t_1 and t_2 and the value $\frac{di_{sq}}{dt}$ represents the backlash.
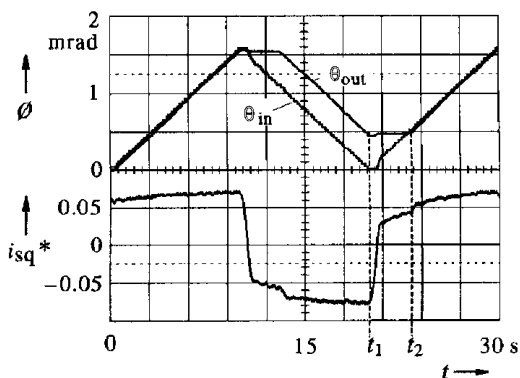
Fig. 7. Principle of backlash identification by a quasi-stationary method

The disadvantage of the method consists in the fact that it requires a smooth course signal in order to ensure precise determination of the contact time, otherwise the method fails.

In ref. [9] there is presented a backlash identification based on motor angular speed sensing ω_M. Fig.8 shows the principle, where $\omega_L(t)$ is the load angular speed, ω^*_M is the angular speed reference signal, $\hat{\omega}_M$ is the maximal motor angular speed. After changing the speed polarity and simultaneously also the current i_q the shaft starts to traverse backlash dead zone at load constant speed: $\omega_L \cong \omega^*_M = const$. The time instant t_2, when there occurs the first contact of gear teeth after traversing the dead zone, presents end of time interval for backlash identification. The value of the backlash 2α can be calculated from the following formula:

$$2\alpha = \left| \int_{t_1}^{t_2} (\omega_M(t) - \omega_L(t))dt \right| \qquad (7)$$
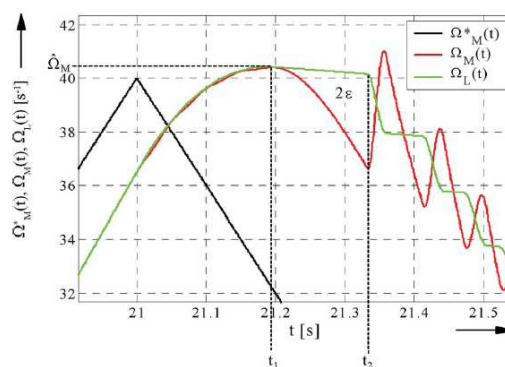
Fig. 8 Principle of backlash identification based on the speed

VII. CONCLUSION

The paper presents an introduction into the topic of precise control of electrical servodrives containing nonlinearities like backlash (for servodrives with gear) and nonlinear friction. Such servodrives are used above all in applications of precise positioning (in case of robots, etc.) and are subject of further research.

REFERENCES

[1] Žalman, M.: *Akčné členy*. Bratislava : Vydavateľstvo STU, 2003. 207s. ISBN 80-227-1835-1.
[2] Wijin Suteeratanapirom, Rapeepong Rattanawaorahirankul, Chutima Laosapan, Peerayot Sanposh, Nattapon Chayopitak, Kanokvate Tungpimolrut, Kunieda, Hiroaki: "*Friction feedforward compensation and position feedforward in CNC machines*," Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), 2010 International Conference on , vol., no., pp.178-182, 19-21 May 2010.
[3] Olsson, H., Astrom, K.J., de Wit, C.C., Gafvert, M., Lischinsky, P.: "Friction Model and Friction Compensation", Eur. J. Control, Vol. 4, No. 3. (1998), pp. 176-195.
[4] Webový slovník, http://www.merriam-webster.com/dictionary/backlash
[5] Nordin, M., Galic, J., Gutman, P. O.: "New models for backlash and gear play. International Journal of Adaptive Control and Signal Processing, vol. 11.(1), pp.49-63 (1997).
[6] Gang Tao; Kokotovic, P.V.: "*Continuous-time adaptive control of systems with unknown backlash* ," Automatic Control, IEEE Transactions on , vol.40, no.6, pp.1083-1087, Jun 19954.
[7] Paško, J., Balara, J.: Model of bearing redactor, *Výrobné Inžinierstvo*, vol. 2 no .8, 2009, ss.36-39, ISSN 1335-7972. In Slovak.
[8] Gebler, D., Holtz, J.: "*Identification and compensation of gear backlash without output position sensor in high-precision servo systems*," Industrial Electronics Society, 1998. IECON '98. Proceedings of the 24th Annual Conference of the IEEE , vol.2, no., pp.662-666 vol.2, 31 Aug.-4 Sep. 1998.
[9] Villwock, S., Pacas, M.: "*Deterministic Method for the Identification of Backlash in the Time Domain*," Industrial Electronics, 2006 IEEE International Symposium on , vol.4, no., pp.3056-3061, 9-13 July 2006.

14

Attractors of the Chua's circuit

[1]*Jakub HARČÁR,* [2]*Milan GUZAN*

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic
[2]Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

[1]jakub.harcar@student.tuke.sk, [2]milan.guzan@tuke.sk

Abstract—This article presents different realizations of Chua´s circuit - realization by two diodes and one operational amplifier with classical and synthetical inductance and by two operational amplifier. The Chua´s circuit can be used in cryptography, signal confidentiality, music, health service and also in automotive technology.

Keywords—chaos, Chua´s circuit, PWL

I. INTRODUCTION

Leon Chua invented the circuit in 1983 to demonstrate chaos in an actual physical model and to prove that the Lorenz double-scroll attractor is chaotic[1]. This electronic circuit is suitable for studying chaos behaviour because its parameters can be precisely controlled and results readily observed by oscilloscope. The circuit is easy to construct using ordinary electronic components like resistors, capacitors, inductors, operational amplifiers and diodes. In 1986, Chua, Komuro, and Matsumoto proved rigorously that the Chua`s attractor is really chaotic.[1]

II. CHUA`S CIRCUIT THEORY

Basic scheme, which is shown in Fig. 1, can be described by three equations, taking into account the voltages and currents shown in figure. These equations are following:

$$C_1 \frac{du_{C1}}{dt} = G(u_{C2} - u_{C1}) - g(u_{C1})$$

$$C_2 \frac{du_{C2}}{dt} = G(u_{C1} - u_{C2}) + i_L$$

$$L \frac{di_L}{dt} = -u_{C2} - R_L i_L$$

where u1 and u2 refers to voltage on the capacitor C_1, C_2 . i_L indicates the current through the coil L and R_L is the resistance of the coil L. Then the function $g(u_{c1})$ is a nonlinear function defined by:

$$g(u_{C_1}) = m_1 v_R + 0{,}5(m_0 - m_1)(|v_R + B_P| - |v_R - B_P|)$$

This function can be displayed as five-segment piece-wise linear (PWL) characteristic shown in Fig. 2 [2]
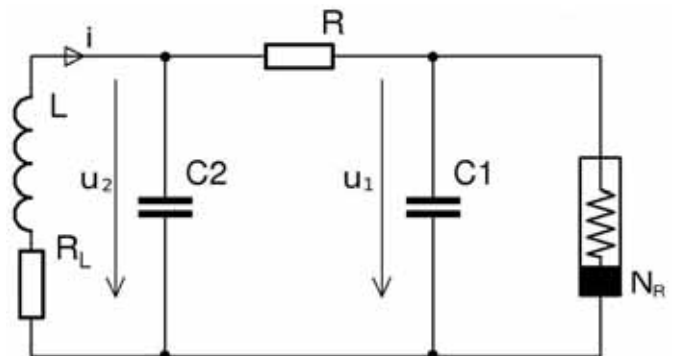
Fig. 1: The basic scheme of Chua´s circuit consisting of a coil, two capacitors, linear and nonlinear resistor
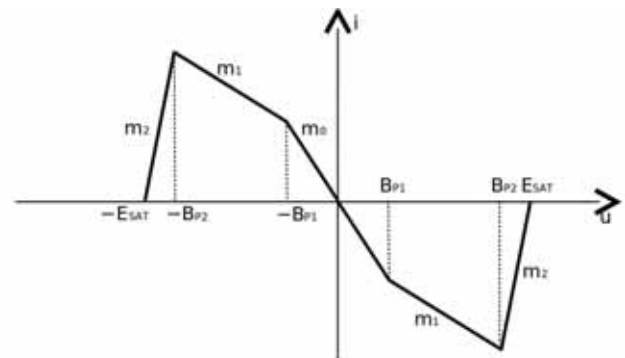
Fig. 2: PWL characteristic of Chua´s diode

Chua´s circuit consists of several basic parts. The base element is a nonlinear resistor (Fig. 1) with PWL characteristic shown in Fig. 2. Other parts are the LC_1 parallel resonant circuit that generates oscillations. Nonlinear resistor is connected with parallel resonant circuit resistor R shown in Figure 1. By change R can be achieved different attractors.

III. REALISATION OF CHUA'S CIRCUIT

In the past, there were several proposals of realization of non-linear resistor of approximately the same characteristics but with different circuit connection. Nonlinear resistor with IV characteristic shown in Fig. 2 is often called Chua´s diode. The first realization of Chua´s diode invented by prof. Chua has been realized through the operational amplifier and two diodes. Prof. Chua himself is inventor of this circuit which is generating chaos.

A. Two operational amplifiers

Kennedy later published in his article [2] the connection using two operational amplifiers. This connection utilizes the
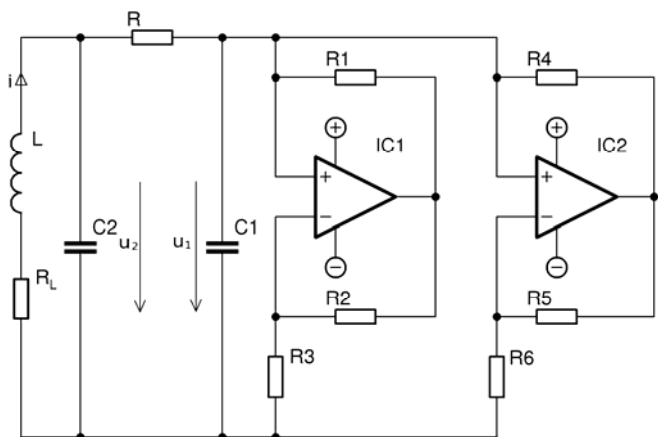
Fig.3 Chua´s circuit consisting of two OA. Values of components: R$_1$=3k85, R$_2$=3k86, R$_3$=613Ω, R$_4$=67.7Ω, R$_5$=67.8Ω, R$_6$=844Ω, C$_1$=4,7n, C$_2$=100n, L=4mH, R$_L$=4,05Ω IC$_1$ a IC$_2$ = MA741.

saturation characteristic of operational amplifiers for the realization of breakpoints in the IV characteristic. Circuit diagram for "double scroll" is shown in Fig. 3. The advantage of this circuit is, that it can work at higher frequencies, but the disadvantage is, that the circuit is sensitive to tolerance of components which affect asymmetries of IV characteristic. Fig. 7 shows the IV characteristic of Chua´s diode.

B. Operational amplifier and a two diode

Chua´s diode can also be constructed through the diodes and operational amplifiers. Position of internal fault determines the diode. Because of the high parasitic capacity of the diodes, the circuit is not able to operate at higher frequencies. The
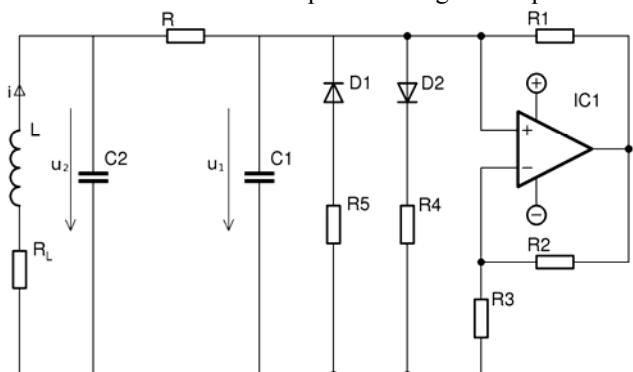
Fig.4 Chua´s circuit consisting of 1 OA and two diodes Values of components: R$_1$=3k85, R$_2$=3k86, R$_3$=717Ω, R$_4$=1261Ω, R$_5$=1261Ω, C$_1$=4,7n, C$_2$=100n, L=4mH, R$_L$=4,05Ω IC$_1$ a IC$_2$ = MA741.

advantage of this circuit lies in easier achievement of circuit´s symmetry which is important for investigation by Rösler´s attractor. Scheme for practical realization of „double scroll" is shown in Fig. 4 [3]

C. Operational amplifier and a two diode with using synthetic coil

Coil´s inductance is the parameter which can affect the behavior of the circuit. By modifying the inductance we are changing the resonant frequency of the parallel pre-resonant circuit. Disadvantages of the conventional coils are mainly their large sizes, no possibility of integration by higher inductance, etc ... As a substitute of conventional coil,
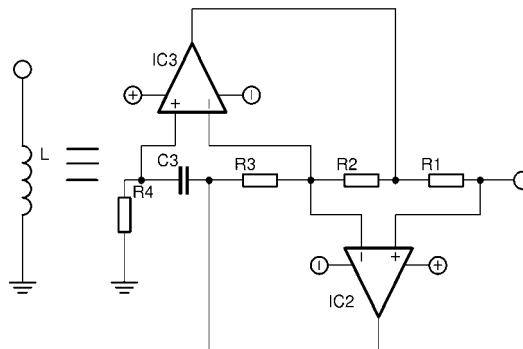
Fig.5 synthetic coil where R$_1$-R$_4$ = 1kΩ, C$_3$=4nF, IC$_2$ a IC$_3$=MA743CN

impedance converter can be used. Its internal connection can be seen in fig.5. Inductance of this impedance converter can be calculated by using the formula: $L = \dfrac{R_1 R_3 R_4 C_3}{R_2}$ [4]

D. Inductance modification can be used in the music industry. This way it is possible to generate tones of a specific timbre. These changes are achieved by the change of C3 in the impedance converter. Large limit cycle san be utilized for tones generation. Components of chaos are not present in a large limit cycle. Deformation of the large limit cycle determines the timbre. Samples of chaos generated by circuits with two operational amplifiers can be found on the web site:
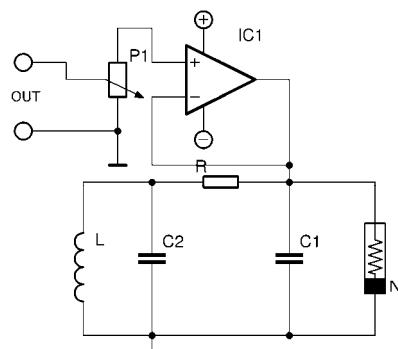
Fig.6 Circuit for connection Chua´s circuit to sound card IC1=MA743CN, P1=10K

http://kteem.fei.tuke.sk/guzan/scyr11/chaos2oz.avi
for operational amplifier and two diode

E. *http://kteem.fei.tuke.sk/guzan/scyr11/chaosoz.avi*
for operational amplifier and two diode with syntetic coil
http://kteem.fei.tuke.sk/guzan/scyr11/chaosozsyntl.avi

IV. MEASUREMENTS

A. Chua´s circuit consisting of 2 operational amplifiers

The circuit was connected like pictured in Fig.3. Resistor R altered in the range of 502 to 692Ω. IV characteristic Chua´s diode together with the change of R is illustrated in Fig. 8. Intersection of Chua´s diode IV characteristic with R forms three unstable singularities P + 0, P-.
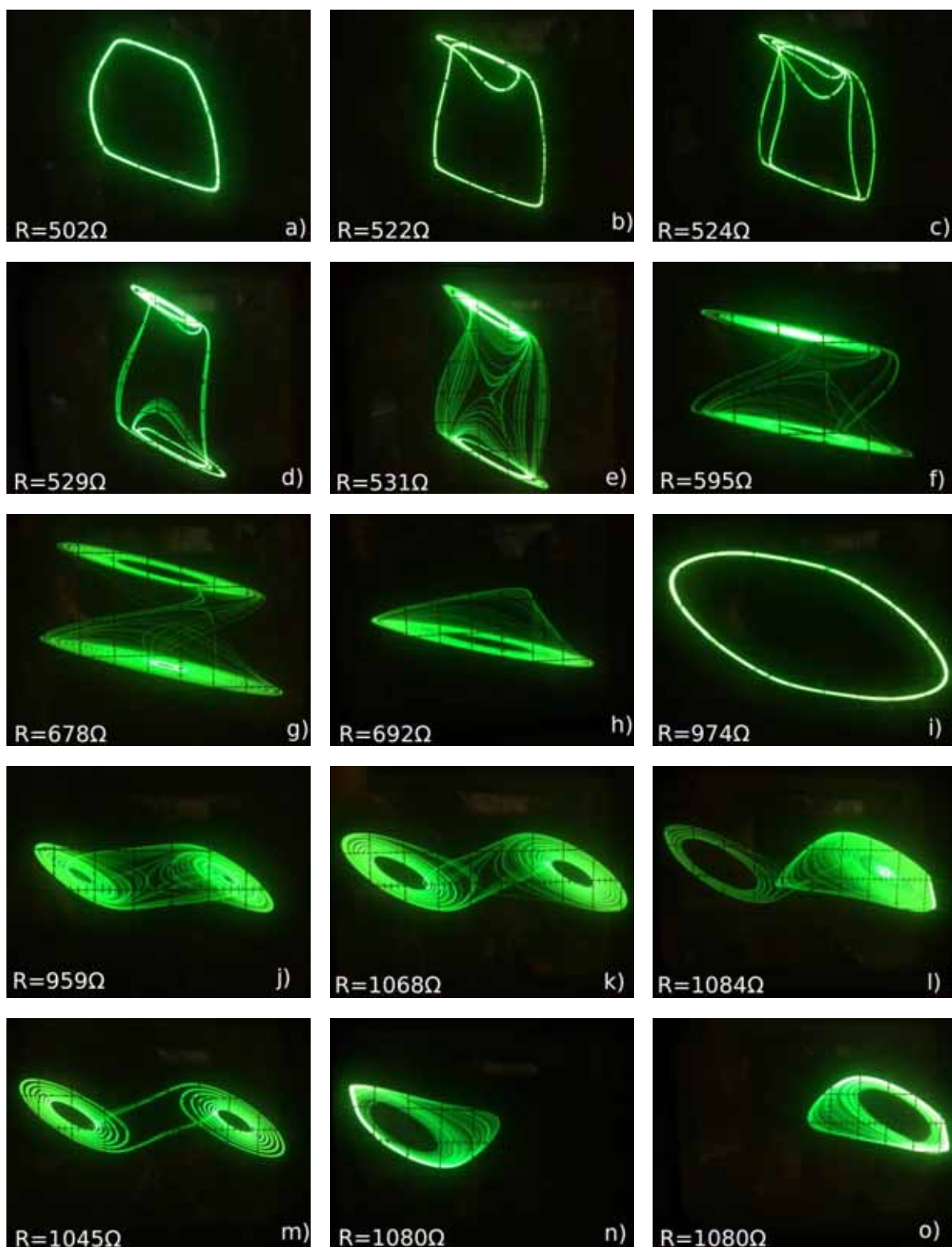
Fig.7 Photographs a - h were taken for connection according to Fig. 3 and phopographs i – o for connection according to Fig.4. For all photographs applies that the sensitivity of input X = 0.5 V. Photos a - e and i – o were taken at oscilloscope´s input sensitivity of Y = 0,2V, photos f - h at input sensitivity of Y = 5V

Fig. 7 to Fig. 7h shows photographs from the oscilloscope for various R. For R = 0 to 501Ω, the circuit was characterized only by one limit cycle (LC) (Fig. 7a).
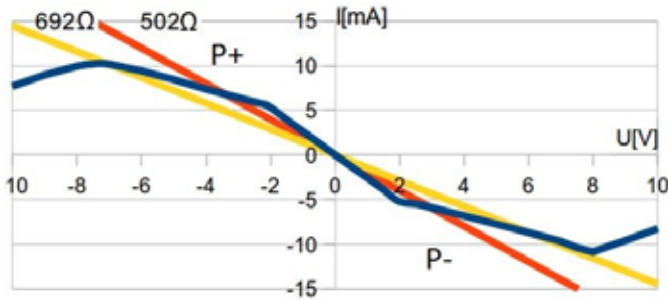
Fig.8 PWL characteristic of connection consisting of two operational amplifiers. The yellow line represents the maximum resistance R in which a chaos arises, and red line indicates a minimum R. When R <500Ω oscillations arose only.

While the value of R was increasing from 502 to 524 Ω the periodic window was observed – Fig. 7b, 7c. It differs from the chaos in Fig. 7d - g in such way, that the displayed signal is periodic. In Fig. 7b, c we can also see the asymmetry of the circuit, which may be caused by tolerances of resistors R_1, R_2, R_4, R_5 as well as unbalance of the operational amplifiers. For R = 529 - 678Ω chaos arises (Fig. 7d-g) around the singularities P + and P . In Fig. 7h, when R = 692 Ω we observe chaos only in the surrounding of the singularity P + or P-. It looks similar to Rösler´s chaotic attractor. This chaos can be switched from neighborhood of P + to P- and vice versa.
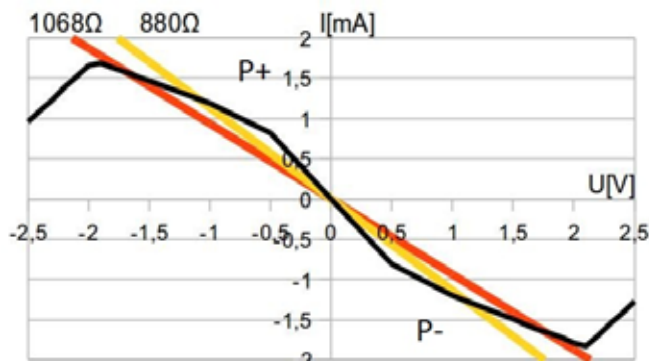
Fig. 9 PWL characteristic with one operational amplifier and two diodes. The red line represents the maximum resistance R in which a chaos arises and yellow line indicates a minimum R

B) Chua´s circuit consisting of 1 OA and 2 diodes

The Chua´s circuit realized in this way together with its parameters is shown in Fig.4. IV characteristic of Chua´s diode and change of R is shown in Figure 9. Similarly as in the Fig. 7a to 7h the similar situation occurred also in Fig. 7i to 7o. In the range 0 - 947 Ω the circuit is characterized by 6m. Both Rösler´s chaotic attractors in the neighborhood of the singularity P + or P- are documented in Fig. 9n and 9o which were together switched to each other by parallel connecting 1µF capacitor to C_2.

C) Chua´s circuit consisting of 1 OA and 2 diodes with using synthetic coil

By using synthetic coil in the connection in Fig.5, the R value
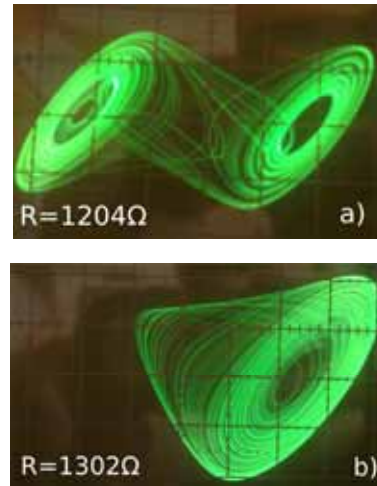
Fig. 10 For both photos applies that the sensitivity of input X = 0.5 V and Y = 0.2V.

was set to range from 950 to 1302Ω. Due to the lack of space, only two photographs are presented. Fig.10 a) represents the chaotic attractor at resistance R = 1204 Ω and the Figure Y shows the Rösler´s attractor at R = 1302 Ω.

V. CONCLUSION

Comparing the results from Fig. 9a - 9h with results from Fig. 9i - 9o leads to observation that the Chua´s circuit in Fig. 4 shows a better symmetry of the chaotic attractors and periodic window than the circuit connected in Fig. 3. Because of this, the circuit in Fig. 4 will be the object of further theoretical and practical examination.

[1]   Mike Audet, "Chaos in chua's circuit," (Unpublished work style)," unpublished.
[2]   Michael Peter Kennedy, " Robust OP Amp Realization of Chua's Circuit" (Electronics Research), California, Berkeley, 1992 pp.66-80.
[3]   Gaurav Gandhi. "Electronic realisation of chaotic circuits" (Unpublished work style)," unpublished.
[4]   Recai Kiliс "A Practical guide for studying Chua´s circuits (Book style)", Singapore: World Scientific Publishing, 2010, pp:17-18

Automobile power supply diagnostic system over the Internet

[1]*Ján MOLNÁR*

[1]Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI, TU of Košice, Slovak Republic

[1]jan.molnar@tuke.sk

Abstract — **A paper is devoted to diagnostic system of automobile power supply based on remote control using Internet. It consists of the detailed description of an automated measurement chain with specific focus on particular parts of the measurement chain. There is also an algorithm of using Internet as measurement bus designed to serve as wireless communication interface. The aim of the paper is also to introduce a specific automated measurement chain designed to diagnostics and evaluation of automobile power supply system.**

Keywords — **automated measurement network, Internet based diagnostics, measurement bus.**

I. AUTOMATED MEASUREMENT CHAIN

General flow diagram of automated measurement chain is displayed on Fig.1. Measurement chain has 1 or more sensors at the input. Using sensor on the input a mutual interaction between measured object (measured parameter of certain natural phenomena or technological process) and measurement system can occur.
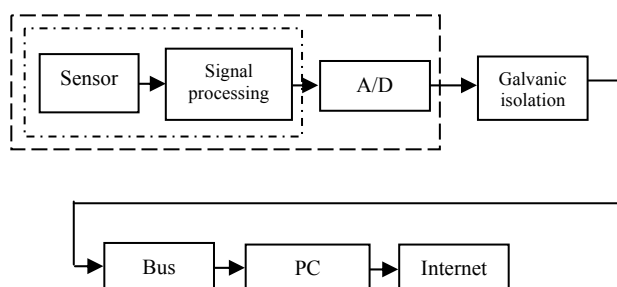
Fig. 1 Automated measurement chain

In the sensor element a measured parameter changes into other physical or chemical quantity. Mostly it is quantity that can be easily transformed into electrical quantity.

Physical or chemical quantity that originates from the interaction of observed (measured) quantity with sensor has only rarely the value or character that can be directly measured or used for further processing. It has to be properly pre-processed and modified in the block for signal processing. This pre-processing includes amplification, filtering, other signal processing, e.g. output impedance, to achieve a source of voltage or current etc.

Processed signal is conveyed on the input of the analog-dialog converter (A/D) where the analog data are transformed into digital code. It is advisable to galvanically separate the further part of the measurement chain due to protect bus.

Then the digital data are sent to bus. Using bus the data are transferred in the digital form to computer where they are processed. These data can be distributed further to distant computer via Internet connection.

The trends in development, e.g. integration and minimizing on of particular machine parts, have brought also changes into measurement chain. Some blocks are integrated into 1 entity or they are eliminated at all. For example sensor and block for signal trimming/processing can be integrated into one united block (it is displayed with dash-and-dot line on Fig. 1), eventually also A/D converter can be implemented into this block additionally (as displayed with dashed line on Fig. 1). The galvanic part is often skipped because new types of data buses are short-circuit-proof. In this way the flow diagram can be simplified with no influence on its functionality. The simplification is only a consequence of implementing new trends into automated measurement chain.

II. WAYS OF APPLICAITON OF INTERNET INTO THE MEASUREMENT PROCESS

There are various architectures that can be applied into measurement chain based on Internet. To choose among the different architecture options, the particular situation and conditions have to be considered.

From connection to network point of view, there are 2 basic possibilities for connecting the device to Internet:

A. Indirect connection – via PC (Fig. 2)
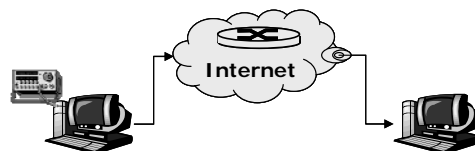
Fig. 2 Indirect connection to Internet

B. Direct connection – without using PC (Fig. 3)

Fig. 3 Direct connection to Internet

In case of indirect connection, a local PC (attached to the measuring device) connected to Internet communicates with

device using different interface (e.g. RS232, LPT or others). Local PC can have function of an interface between remote computer and measuring device or it can fulfil the tasks of a firewall that controls authorization of particular accesses to information, enables coding, compression, filtration or pre-processing of data. It can also serve as a database for measured data. Using computer there can be several devices connected to one node in the same time, whereas computer serves as multiplexer. Indirect connection is preferred when measuring on greater distances.

In the second case it is necessary that the devices include implemented interface for direct connection to network. Nowadays, there are various PLC, oscilloscopes and other multifunctional device equipped by this interface, as well as it is common at intelligent sensors.

The biggest advantage of this kind of connection is a fact that there is no need to connect another PC into the measurement chain. In many of the devices except of the hardware interface there is also software interface implemented, e.g. web server, ftp server, OPC server and others. Direct connection is less demanding if we refer to economic and spatial requirements. Therefore, it is also less demanding for end-user. Direct connection is more suitable for remote measurement in the LAN network (for shorter distances).

From mutual relation of client/server between PC and device point of view, we recognize 2 options:

C. Device as a server (Fig.4)
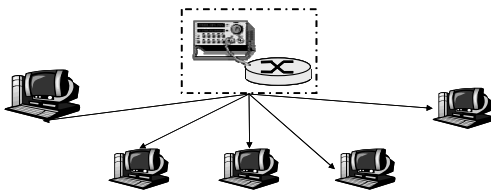
Fig. 4 Devices serves as a server

D. Device as a client (Fig. 5)
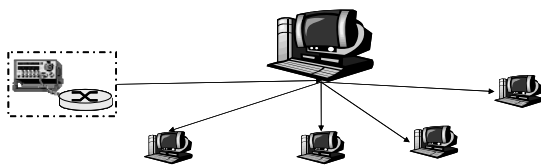
Fig. 5 Device serves as a client

If the device has function of a server, all the connected computers that need the measured data for their activities, connect to the device and then the measured data are sent from device to the computers in the form of multicast messages (from one or more sources the message is sent to several targets). The biggest advantage of this architecture is that powerful server is not necessary in order to provide users the most recently measured data. This type of architecture is suitable for decentralized systems. However, there are few disadvantages. Faster data flow is necessary on the line where the device is connected to higher "intelligence" of the device is needed in order to manage and handle several requests at the same time and the response time can vary according to the number of requests.

If the device has function of a client, measured data are forwarded to the single computer, which will distribute them further according to the need. This type of architecture requires PC in the function of a server. However, there are not very high demands on the bandwidth of the line on which the device is connected as the data are provided for one target only. Therefore, there are relatively constant response times related to this type of architecture. Server provides opportunities for better security and data administration. This type of architecture is more suitable for centralized systems.

III. DESIGN OF AN ARCHITECTURE

Architecture of particular connection is displayed on Fig. 6. It is not focused on the data sampling directly, as this topic will be described in following chapters. It is devoted to „bus" – part of the measuring device that is responsible for transfer of measured data to the device that will assess/evaluate them – to an expert system.
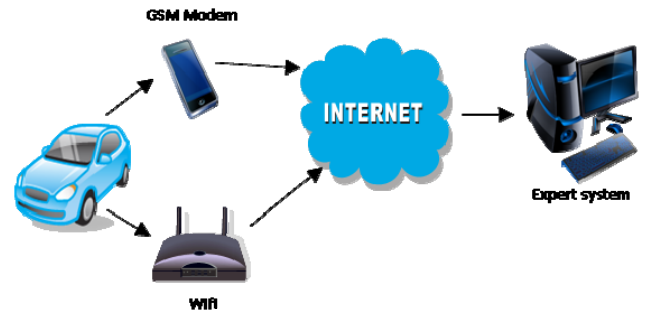
Fig. 6 Design of architecture for distant measurements in automobile

As it is apparent from the figure above, Fig. 6, automobile has embedded GSM modem, device enabling the connection to the wireless network (in our case it is WiFi network). Mobile operator (in our case GSM networks) and access point (for wireless network) enable the connection of automobile to worldwide web - Internet. Using internet connection automobile would be connected to expert system, eventually to computer selected by the producing company.

Measuring system is devoted to diagnostics of power supply of the automobile. Therefore, the daily diagnostic measurement will run with speed of GSM network which is sufficient for this purpose. Entire system is designed to be running daily, for its operation the benefit of the GSM network – wide coverage – can be utilized. WiFi network will be used only during service interventions and for check-up of measuring hardware, or its set-up respectively. It can also be used for creating a back-up of measured data that are stored in the measurement apparatus. WiFi networks of this kind would be built at specialized service workplaces, from which one can very quickly connect to the expert system. Based on the data comparison the database of measured data can be restored in case there are data damaged while data transfer via GSM network. Expert system would be able to restore and renew the missing data and it would achieve complete overview about power supply system of particular automobile.

IV. MEASURED DATA PROCESSING

Measured data sent by measuring setup are displayed after they are received and they are processed and evaluated by expert system made of above mentioned components that saves the data into the database. The data are then processed to further step – data processing. Based on the criteria set by the system, ES can find the disorders, errors and failures.

The record in the table is automatically updated after time interval that can be set by the operator. The table consists of last received data achieved from the measurement apparatus. If there is any discrepancy detected, the special mark is added to the table in order to easily identify its occurrence. The discrepancy does not always equal to failure or error. ES detects the information about switching on and off the automobile. Therefore, table has more columns in order to implement more marks or notes. Fig. 7 provides the example of the table.

			ID:	Napatie	Prud	Datum	Cas	Porucha	Startovanie	Vypnutie
ukaz	ukaz iba prud	ukaz iba napatie	6678913	10.61888	99.973602	2010-06-28	11:19:30		X	
ukaz	ukaz iba prud	ukaz iba napatie	6676707	11.52656	99.973602	2010-06-28	10:42:22		X	
ukaz	ukaz iba prud	ukaz iba napatie	6673059	11.1752	99.973602	2010-06-28	09:40:57		X	
ukaz	ukaz iba prud	ukaz iba napatie	6671338	10.69696	99.973602	2010-06-28	09:11:59		X	
ukaz	ukaz iba prud	ukaz iba napatie	6595261	11.11664	76.793602	2010-06-27	11:11:20		X	
ukaz	ukaz iba prud	ukaz iba napatie	6584171	9.4672	84.211197	2010-06-27	07:54:39		X	
ukaz	ukaz iba prud	ukaz iba napatie	6515130	9.87712	94.947197	2010-06-26	12:12:25		X	
ukaz	ukaz iba prud	ukaz iba napatie	6517779	10.76528	99.973602	2010-06-26	12:57:01		X	
ukaz	ukaz iba prud	ukaz iba napatie	6427807	9.69168	88.896004	2010-06-25	11:32:24		X	
ukaz	ukaz iba prud	ukaz iba napatie	6349355	10.79456	99.973602	2010-06-24	13:01:45		X	
ukaz	ukaz iba prud	ukaz iba napatie	6259331	11.43872	80.502403	2010-06-23	11:36:16		X	
ukaz	ukaz iba prud	ukaz iba napatie	6253813	9.3696	99.973602	2010-06-23	10:03:22		X	

Fig. 7 Graphical representation of voltage and current

As it can be seen at the figure above, table comprises of the columns with information about the voltage, current, date and time. Column with ID is a special identification item according to which one can identify particular record and distinguish it among the others in the database.

Mark „X" indicates where there is a discrepancy identified. If mark X occurs in the column named as "Start", it means that ES has found specific values for voltage and current and identified them at the start and added the identification mark in the particular column. This mark can occur also in the other columns of the table depending on the result how ES evaluates the situation.

It is not easy to imagine the entire situation only from „Start", from information about the voltage and current in certain time. To understand the situation better, the there is a visualization of the situation provided by ES that allows displaying graphically the measured events. Fig. 8 presents the start of the automobile in graphical representation.
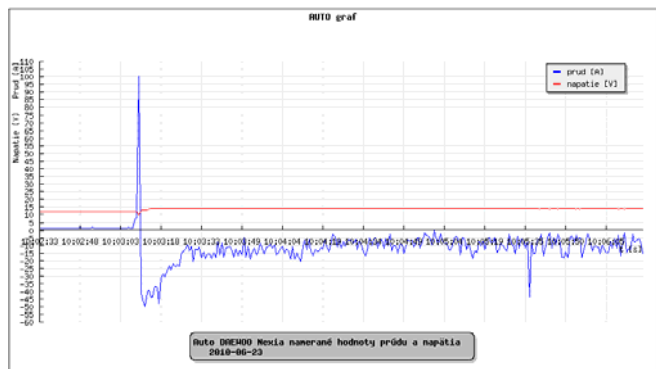
Fig. 8 Graphical representation of automobile's start

Fig. 8 has values of the voltage and current on the y-axis and time on the x-axis. The bottom part of the figure shows the information about the automobile and date of the measurement. Based on these data a general interpretation, a picture about the situation that is happening can be made. Form the figure it can be seen that start of the automobile can be characterized by great expenditures of current from the accumulator and it results to the decrease of the voltage. Based on this information, ES identified the situation as "Start".

The other situation that system is asked to identify is turning of the engine of automobile. Turning off the automobile cannot be as clearly determined as starting the automobile. However, it is necessary to identify also the turning off and it also has characteristic features. Following Fig. 9 shows the turning the automobile off.
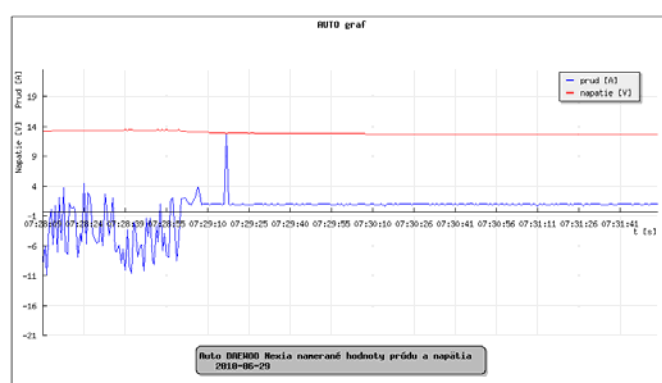
Fig. 9 Graphical representation of turning off the automobile

At the first sight it is visible that the simulation differs from the starting of the automobile. The process of turning off is defined by higher value of voltage at the moment of turning off. In that moment the charging system cannot cover the loss of energy that is spent. ES can identify the turning off the automobile based on this feature.

Fig. 9 showed the standard way of turning off the automobile. Also in this process the failures can occur. They can result in decreasing the lifetime of the accumulator. These failures are caused by systems that are gradually disconnected, e.g. cooling system, alarm, radio etc. Measurements can detect these failures and they can be eliminated by correct timing of switching them off step by step. The previous text, two basic situations characteristic for normal running of the automobile was described. Correct detection of them and subsequent comparison with the recently measured data can help to eliminate and avoid the failures that can occur at starting or turning the automobile off.

Simply said all the other situations excluding those that occur at normal running of the automobile can be classified as failures. However, this is not completely definite. Therefore, the process if setting the criteria for failure identification was the most critical and the most difficult when creating ES. For better understanding, there is situation during running the automobile displayed on Fig 10.
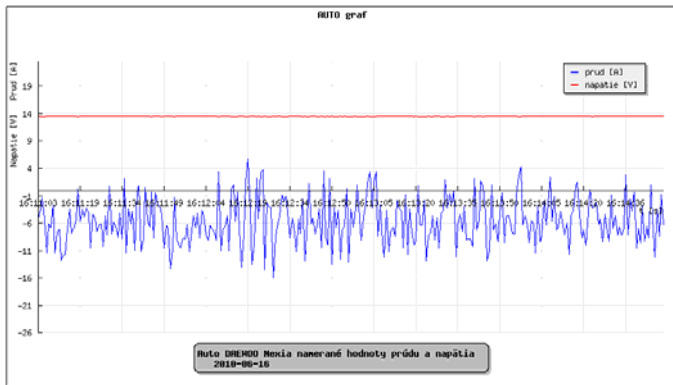
Fig. 10 Graphical representation of running automobile

The values of the running engine are not as definite as they are when starting or turning the automobile off. Therefore, the searching for the failures is more demanding. However, using correct and appropriate setting the criteria of ES, the failures can be identified and marked in the table. Criteria for failure detection must come out of the nature and features of the failure. Mostly it is a short circuit of some electrical device or failure of the charging system. That's how it was possible to determine the conditions for failure evaluation. Or it was sudden decrease of the voltage or it was sudden increase of the consumed current.

Based on these conditions ES can detects and determine the failure when the automobile is running. On the following Fig. 11 there is an easily identified failure of the automobile showed. We can conclude that ES can reliably identify 3 basic situations that can occur when the automobile is running. The other advantage of the ES is that it can browse the measured data and evaluate them based on the criteria set by the personnel. This helps to eliminate the possibility to skip or neglect occurred failure or error that does not match within the criteria boundaries and it to avoid its spreading.
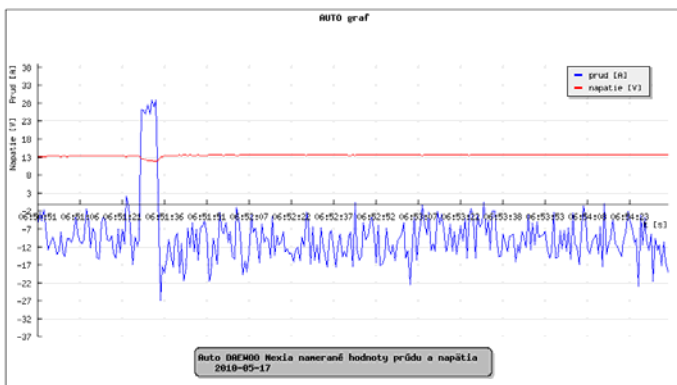
Fig. 11 Failure identified during automobile's running

It is a great benefit that system informs administrator by sending email on defined email account when the failure has occurred. In this manner, the failure can be evaluated,

consulted by qualified professional and conduct the further steps. The failure that was detected by system itself is normally failure that occurred periodically or it is a single-shot error. After the situation is evaluated, the owner of automobile can be contacted and asked to com to check-up.

V. CONCLUSION

Main contribution of the research is an original design and construction of the diagnostic system for power supply of automobile based on the Internet connection for operating the diagnostics. This system in the mentioned design has never been designed or published before. Therefore, it is a unique system. The whole work provides the complex insight in the remote measurements and implementation of new trends into the automotive industry. It provides complete guidelines for the construction of the entire measurement system that cooperates with expert system for evaluation. Altogether these two systems create a complex diagnostic and measurement system.

REFERENCES

[1] Kováč, D., Kováčová, I., Molnár, J.: Elektromagnetická Kompatibilita - meranie, Technická univerzita v Košiciach, Košice 2008, 1. vydanie ISBN 978-80-553-0151-8

[2] Kováč, D., Vince, T. , Molnár, J.: Modern Internet Based Production Technology. In: New Trends in Technologies, SCIYO Publisher, Croatia, 2010, 20 pages, ISBN 978-953-307-212-8

[3] Vince T., Kováčová I.: Distance control of mechatronic systems via Internet, In: Acta Electrotechnica et Informatica, 2007, Roč. 7, č. 3 (2007), s. 63-68. - ISSN 1335-8243

[4] Molnár, J., Vince, T.,: Telemetric expert system based on interent. In: AMTEE '09, September 7.-9.9.2009, Cheb, Czech Republic. p. 1-4.ISBN 978-80-7043-821-3.

[5] Tomčíková, I., Molnár, J., Vince, T.: Interakcia magnetického poľa a napätosti pre elastomagnetický snímač tlakovej sily. In: Elektrorevue. no. 34 (2008), p. 34-1-34-11

[6] Vince, T., Kováč, D: Remote and regulation of electrotechnical systems via Internet (in Slovak). Publisher: FEI TU Košice 2010, 1st edition, 130 pages, ISBN 978-80-553-0571-4

Bi-directional DC/DC Converter for EHV

Tomáš Béreš, Marcel Bodor, Erik Eötvös

Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

tomas.beres@tuke.sk, marcel.bodor@tuke.sk, erik.eotvos@tuke.sk

Abstract— **Bi-directional buck-boost DC/DC converter for hybrid battery in electric hybrid vehicle (EHV) controlled by UC3637 is described in this paper. The first part of the paper is aimed at concept of hybrid battery. Design of power circuit and control circuit with UC3637 of converter is described in the second part of the paper. Experimental results from measuring of converter are mentioned in last part.**

Keywords— **Converter control, hybrid battery, pulse width modulation (PWM), electric hybrid vehicle (EHV)**

I. INTRODUCTION

The last years are characterized by rapid development of electronic systems, which uses an accumulator as a basic power supply. However, presently the accumulators are the weakest element of the power electronic supply system. It is caused by low dynamics of input power, temperature dependence, short lifetime and a lot of other limitations. The most significant improvement in recent 200 years has been achieved by developing ultra capacitor (UCAP). The ultracapacitor has much better electrical parameters than conventional accumulator. The next table shows comparison of the features of ultra-capacitor, accumulator and classic capacitor.

Available Performance	Accumulator	Ultra – capacitor	Classic capacitor
Charge Time	1 – 5 hrs	0,3 – 30 s	$10^{-3} – 10^{-6}$s
Discharge Time	0.3 – 3 hrs	0,3 – 30 s	$10^{-3} – 10^{-6}$s
Energy (Wh/kg)	10 - 100	1 – 10	<0,1
Cycle life	1000	10^6	-
Specific Power	< 1000	10 000	>100 000
Charge/Discharge Efficiency	0.7 – 0.85	0.85 – 0.98	> 0.95

Tab.1. Parameter comparison of ultra-capacitor with accumulator and classic capacitor

At present the low energy density is main disadvantage of ultra-capacitors. One of the possibilities is to fuse the advantages of ultra-capacitors and high energy density of accumulators to a hybrid secondary power source.

II. CONCEPT OF HYBRID BATTERY

Hybrid battery (HB) is a name for an improved topology of secondary voltage power source. Its output power dynamics and lifetime considerably exceed the recent types of accumulators. The hybrid battery is in nature a cascade connection of an ordinary accumulator with an ultracapacitor via a bi-directional DC/DC converter as it is seen in Fig. 1.
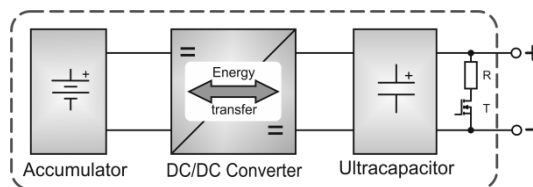
Fig. 1. Block diagram of hybrid battery

High dynamics of input-output power of the hybrid battery is achieved due to the ultra-capacitor. It means that high dynamic parameters of the hybrid battery are given by the parameters of the ultra-capacitor and static parameters by the accumulator. Bi-directional DC/DC converter is a main part of a hybrid battery. The converter has essential influence on the operational properties and the efficiency. Recuperation conditions of the bi-directional DC/DC converter are given by the use of an accumulator in hybrid battery.

III. DESIGN OF DC/DC CONVERTER

The parameters of proposed DC/DC converter are shown in Table 2.

Parameter	Value
Input voltage Uin	**15-30 V**
Output voltage Uout	**14/24 V**
Max. output voltage ripple ΔUout	**5%**
Max. output current	**10A**
Max. current ripple	**1A**
Switching frequency	**50kHz**
Efficiency	**>80%**

Tab. 2. Table of parameters

A. Power circuit of DC/DC converter

Power circuits of the DC/DC converter are in the Figure 2.
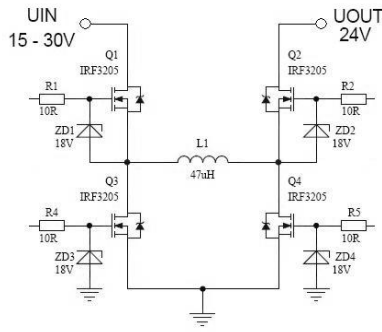
Fig.2.Topology of bi-directional buck-boost DC/DC converter

The bi-directional converter consists of two buck-boost converters connected in cascade. This converters are interconnected through inductance i.e. boost converter with buck converter (Fig.2).

The value of output voltage in general is:

$$V_{OUT} = V_{IN} \frac{D_1}{D_2} \qquad (1)$$

where:

$$D_1 = \frac{t_{Q1(ON)}}{T} ; \qquad D_2 = \frac{t_{Q2(ON)}}{T} \qquad (2)$$

$t_{Q1(ON)}$ and $t_{Q2(ON)}$ indicate the ON time of the MOSFET switches Q_1 and Q_2 respectively, whereas T is the switching period.

Cascaded buck-boost converter can work in three operation modes, which will be introduced below.

a) Buck mode($V_{IN} > V_{OUT}$)

Transistor Q_2 is always ON and Q_4 is always OFF during this mode ($D_2 = 1$). Only Q_1 and Q_3 are switching synchronously. In this operation mode the cascaded buck-boost converter works as classic buck converter. Then the value of V_{OUT} is for buck mode as follows:

$$V_{OUT} = V_{IN}.D_1 \qquad (3)$$

b) Buck-boost mode ($V_{IN} \approx V_{OUT}$)

In this switching mode all four MOSFET's operate during the period. The first path (Q_1, Q_4 are ON) enables charging the inductor, the second path (Q_2, Q_3 are ON) allows the energy stored in the inductor to be delivered to the output capacitor. This way of switching determines the following relation between D_1 and D_2:

$$D_2 = 1 - D_1 \qquad (4)$$

By combination of the equations (1) and (4), the following expression is obtained:

$$V_{OUT} = V_{IN}.\frac{D_1}{1 - D_1} \qquad (5)$$

c) Boost mode ($V_{IN} < V_{OUT}$)

Transistor Q_1 is always ON and Q_3 is always OFF during the period in this mode ($D_1 = 1$). Only Q_2 and Q_4 are switching synchronously. In this operation mode the cascaded buck-boost converter works as classic boost converter. Then the value of V_{OUT} is for boost mode as follows:

$$V_{OUT} = \frac{V_{IN}}{D_2} \qquad (6)$$

B. Control circuit with UC3637

Switched mode controller UC3637 is used for control of this converter. Scheme of control circuit is shown in Figure 3.
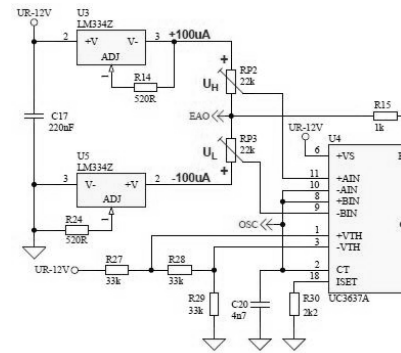
Fig. 3. Control circuit with UC3637

Amplitude of triangle waveform oscillator (+U_{TH}; -U_{TH}) is set by voltage divider R15, R17, and R20. Value U_H and U_L is set by trimmer RP2 and RP3. Changing modes of converter depend on value U_H and U_L as follows:

If: U_{EAO}-U_L < U_{-VTH} and U_{EAO}+U_H < U_{+VTH}
buck mode is set
If: U_{EAO}-U_L > U_{-VTH} and U_{EAO}+U_H < U_{+VTH}
buck-boost mode is set
If: U_{EAO}-U_L > U_{-VTH} and U_{EAO}+U_H > U_{+VTH}
boost mode is set
where U_{EAO} is output from voltage PI regulator.

For better understanding the function of controller is shown in Figure 4.
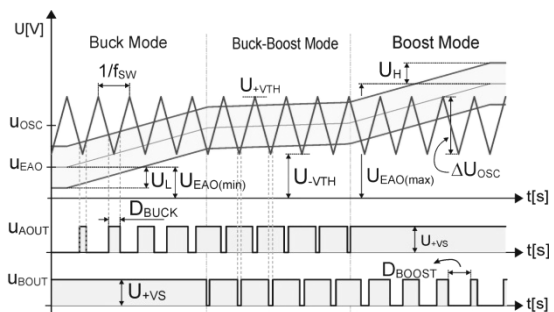
Fig. 4. Function of control circuit

Voltage u_{AOUT} is input for transistors Q_1 and Q_3 driver, voltage u_{BOUT} is input for transistors Q_2 and Q_4 driver. Transistors Q_1 and Q_2 are switched by non-inverted signal and transistors Q_3 and Q_4 are switched by inverted signal.

IV. EXPERIMENTAL RESULTS

The function of the proposed DC/DC converter was verified on the laboratory model. Principle of signal creation with controller UC3637 for drivers is displayed on the following oscillograms.

The first oscillogram (Fig.5) was captured in buck mode of the converter. Transistors Q_1, Q_3 are switched. Transistor Q_2 is always ON, Q_4 is always OFF in this mode. The second oscillogram (Fig.6) shows buck-boost mode operation of the converter. In this mode all transistors are switched in diagonal pairs.
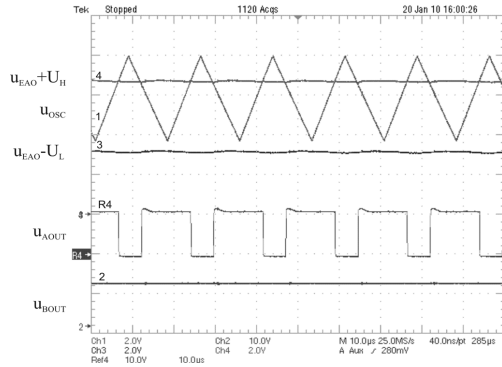
Fig. 5. Control signal creation with UC3637 for drivers in buck mode (U_{IN}=28V,U_{OUT}=24V)
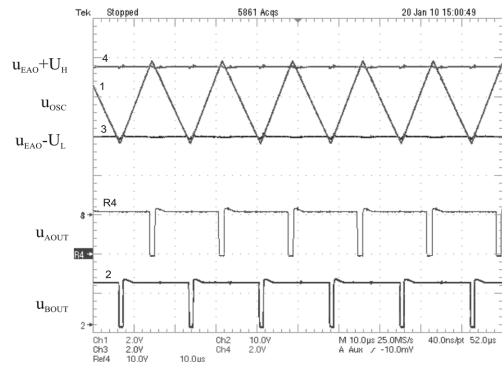
Fig. 6 Control signal creation with UC3637 for drivers in buck-boost mode (U_{IN}=25V, U_{OUT}=24V)

Boost mode operation of the converter is displayed in the third oscillogram (Fig.7). Transistors Q_2, Q_4 are switched. Transistor Q_1 is always ON and transistor Q_3 OFF.
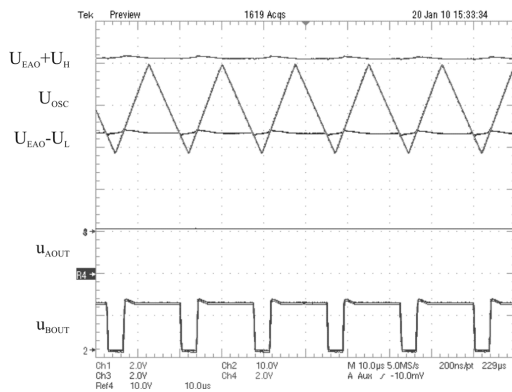
Fig. 7. Control signal creation with UC3637 for drivers in boost mode (U_{IN}=18V, U_{OUT}=24V)

Voltage of transistors Q_1-Q_4 and inductor current i_L in buck-boost mode is shown in Fig.8. Correct function of

converter control is shown in this oscillogram. All four transistors are switched on simultaneously.
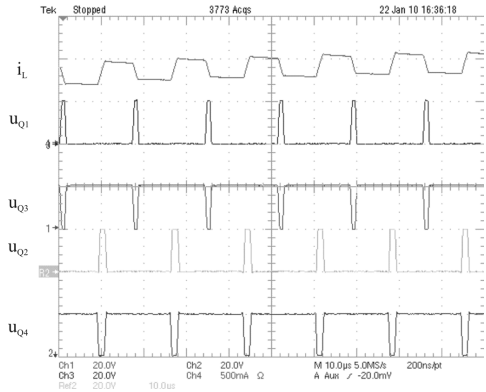
Fig. 8. Voltage of transistors Q_1-Q_4 and current of inductor in buck-boost mode (U_{IN}=20V, U_{OUT}=20V, I_{OUT}=0,6A)

Experimental model of DC/DC converter is shown in Fig.9.
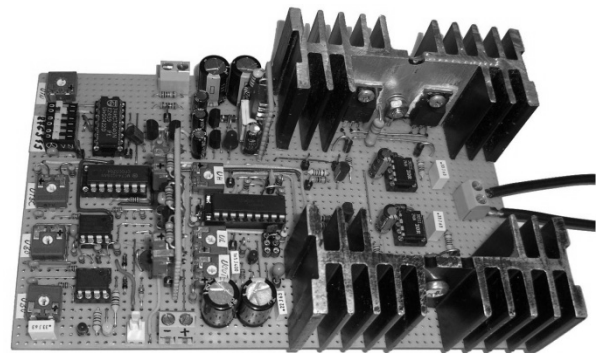
Fig. 9. Photo of DC/DC converter

V. CONCLUSION

New concept of control method for bi-direction buck-boost DC/DC converter is described in the paper. This concept of control decreases power transistor switching losses and thus increases efficiency of converter.

REFERENCES

[1] Gaboriault M., Notman A.: A High Efficiency, Noninverting, Buck-Boost DC-DC Converter, APEC '04. Nineteenth Annual IEEE Volume 3, 2004

[2] Markel T., Zolot M., Sprik S.: Ultracapacitors and Batteries in Hybrid Vehicles, National Renewable Energy Laboratory (NREL) PR-540-38484, 8.2005, pp.7-15.

[3] J. Hamar, I. Nagy, P. Stumpf H. Ohsaki, E. Masada: New Dual Channel Quasi Resonant DC-DC Converter Topologies for Distributed Energy Utilization

[4] Dudrik J.: High Frequency Soft Switching DC-DC Power Converters. Monograph, ELFA, Košice 2007, ISBN 978-80-8086-055-4

[5] Tereň A., Feňo I. Špánik P.: DC/DC Converters with Soft (ZVS) Switching. Proc. of the Int. Conf. ELEKTRO 2001, section - Electrical Engineering. Žilina 2001, str.82 – 90.

[6] Texas Instruments: Datasheet UC3637 Switched Mode Controller for DC Motor Drive, www.ti.com

14

Bridge rectifier controlled by PLC

Michal Števák

Supervisor of bachelor work: prof. Ing. Dobroslav Kováč, CSc.

Department of Theoretical Electrical Engineering and Electrical Measurement FEI TU of Košice, Slovak Republic

michal.stevak@student.tuke.sk

Abstract - **This paper describes the use of PLC (Programmable Logic Controller) S7-300 from Siemens. With it is possible to control one-phase impulse-controlled rectifier in bridge connection and its possible practical use in industry for welding metals DC electrical current.**

Keywords - **PLC, Simatic, controlled bridge rectifier, welding**

I. INTRODUCTION

PLC - Programmable Logic Controller is a technical device (basically a programmable digital computer), which is used to automate management processes. Currently the most widely used PLC control system, which were applied, in different areas of automated techniques in the management of machines and technological processes in various industries.

II. PLC TYPE OF SIMATIC S7 300

Siemens AG is one of the leading electrotechnic groups in the world. It covers virtually entire spectrum of electronics and electrotechnics. In the field of automation control equipment is a leader in the European market. It offers a wide range of products for industrial automation.

Type of Simatic S7 300 controllers is one of the most successful in the market. It offers high performance for construction machinery and equipment, not only in the engineering industry. It is characterized by high stability at work with little maintenance and contains a number of integrated functions.

Fig. 1. CPU 313C

In this work is used programmable logic adapter from the type Simatic S7 300, namely the CPU 313C (its structural design is in Fig. 1). It is a compact adapter that in addition to CPU and power supply already contains input and output modules.

The input module offers the possibility to connect 24 digital and 5 analog inputs, 16 digital output module and 2 analog outputs. Furthermore, it contains technology functions including reading, measuring the frequency and PID controller. Program can be backed up by Micro Memory Card. Communication with the PC interface is provided by MPI It can be modularly extended a wide range of other functional modules for individual use and automated management tasks.

III. PRINCIPLE OF CONTROLLED RECTIFIER

Rectifiers consist of semiconductor devices are currently the most widespread type of converters. Collected energy from the AC mains and used to convert AC electricity to DC.

The basic structure of the controlled rectifier is show in Fig. 2. It consists of power supply transformer (Pt), power circuit (PC) and a control circuit (CC). In this case it is PLC control circuit, which generates control signals synchronized with line voltage and variable phase shift (phase control power output). These signals are fed to the control electrodes of thyristors. Signals are used to control voltage and rectified current of rectifier and also connected by a controlled object.

In connection with phase control concept is used steering angle. This term means shifting impulse control, so the time when the thyristor being locked against the time when the line voltage commutation occurs naturally.
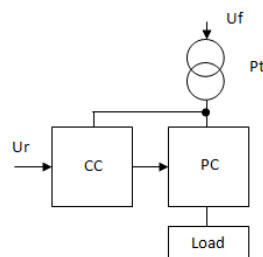
Fig. 2. The basic structure of the controlled rectifier

IV. PRACTICAL IMPLEMENTATION OF MANAGEMENT

A. Controlled power of the involvement

In this case, the controlled circuit of two-phase pulse

controlled bridge rectifier. It is powered by one stage and two-pulse term means that, over one period of supply voltage. The number of dissenting commutation, it means the number of peaks acquired time course of the output voltage for a single period equal to two. This is a rectifier with natural commutation (commutation networking). It means that the current transition from one pair of thyristors to the second is due to voltage drop on the stage, which results in an increase in electrical current and voltage phase, which has a current download.

Rectifier wiring diagram is in Fig. 3. Rectifier is supplied from the AC mains voltage of 230V. Control electrodes of thyristors are connected to digital outputs of the PLC, which serves as a pulse generator for the thyristors. The controlled rectifier is used bridged foursome thyristors BT151-800R. Thyristors are in plastic TO220 for maximum voltage 800V and manage constant load electric current to 12A.

Thyristors TY1 and Ty4 are connected to a common digital output DO1 adapter CPU 313C and thyristors Ty2 and Ty3 of common digital output DO2 adapter CPU 313C. The value of the logical unit for output voltage 24V PLC responds. Pulses to control electrodes thyristors are therefore introduced via voltage dividers (maximum voltage in thyristor control electrode is 5V). PLC output voltages are divided voltage dividers, which are a combination of 4.7 k ohms resistors and 910Ω to control the electrode voltage reaches about 3.9 V, which is for the closure is fully sufficient. For a bridged rectifier output is then connected load, which power management, is required.
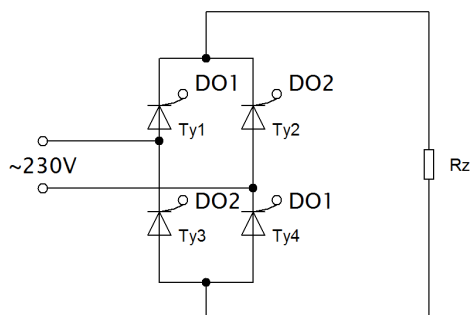
Fig. 3. Schematic of two-phase controlled rectifier impulse in bridged

B. Management of the involvement

The PLC is composed of CPU (central processing unit), which is running firmware (operating system). Its task is the proper functioning of the PLC, communication with the modules and run user program). PLC works with both input and output signals based on image process. Images of inputs and outputs are stored in data memory. User-generated program then works with these images and does not proceed directly to the input and output modules of PLC. Processing and image update process automatically makes the processing unit in a certain sequence. This sequence is called operating cycle of the processor. Its sequence for this case management is shown in Fig. 4.

1. The first step is loaded images of signals from analog input module (AI1), which are stored in data memory
2. Program will control implement a bridged rectifier
3. After its completion, the result of the program recorded in the output process image data in memory
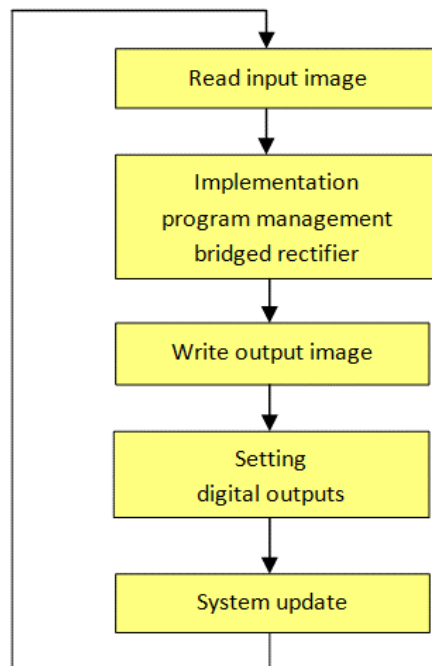
Fig. 4. Operating cycle of the processor

4. Output process images output from data memory is entered on the digital output modules of PLC (DO1 and DO2)
5. Finally is system and update state variables and preparing for the new cycle

The duration of one operational cycle is of the order in units of up to hundreds milliseconds and is dependent on the complexity of the user program as well as the quantity of connected inputs and outputs. In this case is used a single analog input and two digital outputs it is a time exceeding a little over 1 to 1.5 ms.

Block diagram for control adapter CPU313C bridged rectifier is shown in Fig. 5. Mains voltage is transformed to a value in the range of values that can handle PLC input module. This value is channelled bridge rectifier and fed to the analog input adapter CPU313C. Analog input module allows for the processing the AC signal in the range of ± 10V. To control rectifier is sufficient only work with the absolute value of the input signal (controlled by the same power is part of the guidelines half-wave positive and negative). For this reason is signal fed through the input rectifier bridge and works only with positive values of the input signal. Analog input module is set to a voltage range of 0 to 10V and suppresses interference frequency to a value other than 50Hz (400Hz).
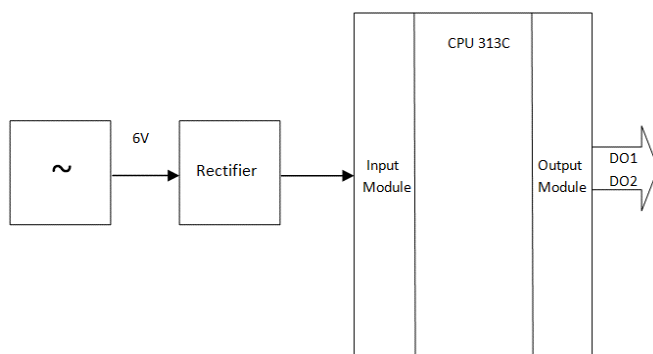
Fig. 5. Block diagram for control adapter CPU313C bridged rectifier

In the analog input module is used 12-bit analog / digital converter. PLC memory can store data on the size of 1 byte (8 bits), respectively 1word (16 bits). Therefore, the output of the converter is stored in memory as one word and 4 bits of lower order are neglected and thus equal to 0. The highest bit is used as a sign to distinguish positive and negative values. In this case, to work only with positive values is still equal to 0. The speed with which the converter is able to convert one sample from analog to digital values is in the range around 1 to 1.5 ms. One half-period AC voltage with a frequency of 50Hz can be divided into 7 to 10 samples.
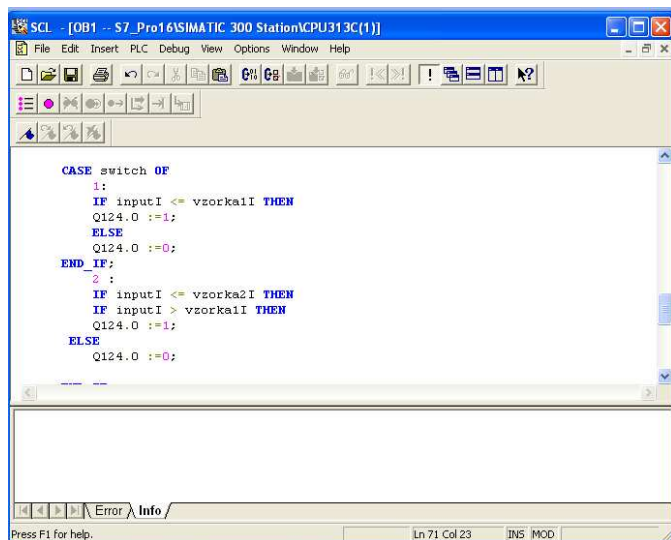
However, it is necessary to consider only the case with the smallest number of samples. Seven instantaneous values of voltage can be obtained for one half period.

Created user program works by comparing the measured instantaneous voltage values against the reference. After calculating the reference value regulation is divided into 7 equal power levels.

An important point in management is that the measured instantaneous and reference value will never match (for a period of one half period is not possible to retrieve the same number of samples). Therefore, the program is built so as to identify in which interval the measured value is immediate. Whereas reference values are the same in the first half period and in the second half, another problem is how they can be differentiate. The program is treated so that it is detected when the half-cycle signal into a maximum value. It is recorded with a logical unit in the assistance variable and it is obvious in which half is the value.

If the current value within the range of values is set to digital output DO1 brought the value of the logical first. Then are switched thyristors TY1 and Ty4. The same also happens in other parts of half period. Digital output DO2 is set to logic 1 and is switch pair of thyristors Ty2 and Ty3.

Program management adapter CPU313C was created in the language of S7-SCL, which is part of the software package. Part of the source code is shown in Figure 6.

Technical options in CPU313C adapter allows control only in seven degrees of power. This involvement would not be appropriate for practical use. It only serving to demonstrate one of the many uses for automated PLC control. For practical use would be appropriate to use a powerful PLC for control that might allow more precise output regulation. Another option would be used by microprocessor control. In this case it would probably be a better solution uses a microprocessor because powerful PLC purchase would be financially quite difficult.

References

[1] Borba, L.: Electrical drives, 1st ed., STU Bratislava, pp. 57-61
[2] Šmejkal, L. - Martinásková, M.: *PLC and automatization 1st part – Basic terms, introducing into programming*, 1st ed., BEN-technická literatura Prague, 1999, pp. 49-51
[3] SIMATIC Controllers, [online]. Available on the Internet: <https://www.cee.siemens.com/web/slovakia/sk/corporate/portal/produkty/divizie/automatizacna/ponuka/priemyselne/Pages/simatic_controllers.aspx>
[4] S7-300 Starterkit 2009 – Get off to on easy start with SIMATIC, [online]. Available on the Internet: <http://www.automation.siemens.com/mcms/programmable-logic-controller/en/simatic-s7-controller/s7-300/starterkit/Pages/Default.aspx>
[5] PLC and their programming - 1. What it is PLC?, Available on the internet: <http://www.posterus.sk/?p=6903>

Fig. 6. Part of source code programmed in the language of SCL

V. PRACTICAL USE CONNECTIONS

When used in a controlled thyristors bridged rectifier that handles higher electrical currents and suitable welding transformer could be used as the engagement model welding unit designed for soldering of direct electrical current.

Calculation of regulators of vector control system taking into account nonlinearity of the magnetization curve

[1]Anton SVISTUN, [2]Vita OGAR, [3]Ján MOLNÁR,[4]Martin BAČKO

[1,2]Dept. of Automatic Control Systems and Drives, Kremenchuk Mykhaylo Ostrohradskyi National University, Ukraine
[3,4]Dept. of Theoretical Electrical Engineering and Electrical Measurement, Technical University of Košice, Slovakia

[1]profi268@i.ua, [2]vogar@kdu.edu.ua, [3]jan.molnar@tuke.sk, [4]martin.backo@tuke.sk

Abstract—**In this paper there is presented an unused before method of accounting a nonlinearity of magnetization curve in vector control systems (VCS) with the field orientation by the rotor flux. Also there is presented additional calculation for regulators of current, flux and angular speed. Finally, shown later on figures with the charts of transients, VOC is very perspective in application in high-accuracy, fast-response systems.**

Keywords—**Vector Control Systems, magnetization curve, rotor flux, current regulators.**

I. INTRODUCTION

Development of technical systems, an appearance of microprocessors and related research papers in the field of electric drive, allow to apply an AC electric drive, where it would seem possible to apply the DC Drive. DC drive has excellent adjustment characteristics, hereupon DC motors are widely applicable in manufacturing. AC drive has greater structural reliability, simplicity of construction and lower price compared with DC drive. Low price and high reliability makes AC drive undeniable competitor of the DC drive. In control and automation systems appliance of the AC drive is connected with certain difficulties. In terms of automatic control, its quite difficult to operate AC drive and measure its coordinate gages. For example, angular velocity of shafting depends on supply frequency, and on shaft load, which results in complicating of control system. During creating projects based on AC drives, there always emerges a need of providing high dynamic parameters in transient states, high-accuracy of maintaining the rotor speed, and thus, high energy efficiency rates. For those purposes a vector control system is used.

II. PRINCIPLE OF OPERATION

On the method of vector control is written pretty much literature has resulted: block diagrams, methods of synthesis of regulators, the description of coordinate conversion, and more. In practice, implementation a vector control system in simulation program is quite difficult, without depth of knowledge about AC drive. For those purposes there is not much information in [1-4,5] and example of vector control system design with rotor flux orientation, could be repeated without assistance. Examples of regulators synthesis, which could be used for creating and developing control system, within structure shown in [1, 2], are given in [3] and [4]. An example given in [5] is not enough to be rebuild by oneself, because of presenting with only one block, with is not given an example with description. Regulator synthesis shown in [3] is fit with structures, given in [1], [2].

With the α axis orientation on the rotor flux linkage, an analogue between control systems of AC drive and DC drive can be done, which could be seen on fig.1 and fig.2. This allows making a conclusion about axis operating within AC drive similar to DC drive.
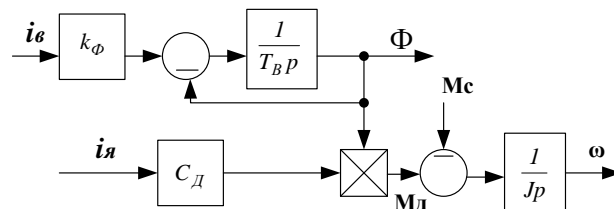
Fig. 1 – Block diagram of DC drive: iв – excitation current; iя – armature current; Tв – exciting coil time constant; $к$ф – constructive coefficient; $с$д – engine constant; J – moment of inertia; Mд – torque; Mс – drag torque.
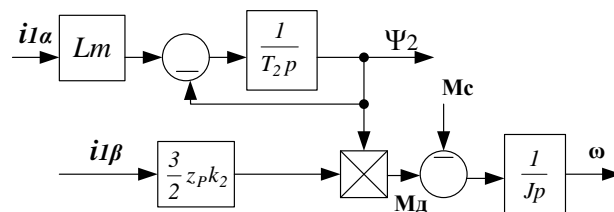
Fig.2 – Block diagram of induction motor:
$i1\alpha$, $i1\beta$ – reactive and active components of stator current; $T2$– electromagnetic constant; $k2$–coefficient of electromagnetic rotor connection; zp – number of pole pairs.

Comparison of control systems – scalar and vector, which gives a conception about obvious advantages of vector system is presented in Tab. 1.

TABLE I
COMPARSION CHARRACTERISTICS OF CONTROL SYSTEMS

Features	Vector system		Scalar system	
	Without speed gage	Icnremental encoder	Without speed gage	Icnremental encoder
Accuracy of regulating	0,2%	0,02%	20%	10%
Regulating width	1:100	1:1000	(6-10):1	≈2/sk
Method of regulating	Ψ2= const		U/f=cons	
Overloading ability	Constant at all frequency values		Poor at the low values of the frequency	
Table shortages: sk – critical slip of the electric drive, Ψ2 – rotor flux linkage, U – stator voltage, f – frequency of the stator voltage.				

A view of AC drive in [6] is most widely used for analyses and synthesis of electromagnetic process in electro drives with vector control. However, if flux linkage under work of AC drive has significant changes, then the desired accounting of modifying parameters of mathematical model is related with effect of saturation of magnetic system. For AC drives, which are supplied from frequency converter, it is usually assumed that, the multiplicity of output current in operating mode of electric drive is limited (of not more than 2-3 values of rated current) and the saturation effect of circle scattering can be neglected. This is confirmed in [7]. Much more significantly affects the effect of saturation of magnetization circle [3]. In particular, in operating mode AC drive from 4A series, could have a magnetization inductance Lm varying up to 30% [8,9]. In AC drive flux linkage adjustment required in the following modes:

1) at work over at speeds in excess of nominal (in the 2-nd zone of speed control at a constant power, with field weakening);

2) while optimizing energy performance, over the need to regulate the flow of magnetization, depending on the load;

3) while optimizing the dynamic characteristics of the drive, including processes performance.

To account the effect of saturation, a method of static inductances is used static, which gives, however, very high accuracy, while describing the dynamic processes. In this method taking into account the nonlinearity of magnetization curve dependence with static dependence between flux linkage and excitation current, that is set up by table of values, function or analytical approximations.

III. LABORATORY MODEL OF THE VECTOR CONTROL SYSTEM

On materials of the researches presented in [3,4,10,11], on the basis of a t-shaped equivalent circuit, taking into account nonlinearity of the magnetization curve (рис.3), it is necessary

to calculate a vector control system (in α-β-0 coordinates) [10] for the AC drive type 4A100L4У3, with following passport data (Tab. 2).
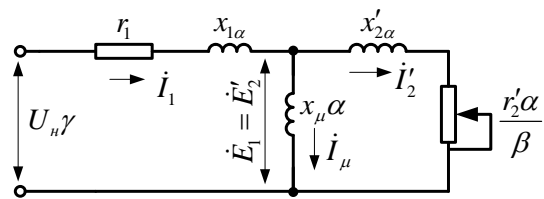
Fig. 3 – T-shaped equivalent circuit of AC drive with frequency control.

TABLE II
Passport data for AC drive 4A100L4У3.

Ωn,s^{-1}	150,5	Xs, Ohm	2,28
Mn, N·m	26,6	Xr, Ohm	3,15
Ms, N·m	55,8	Xm, Ohm	57,4
Mcr,N·m	58,48	Ls, H	0,19
In, A	8,61	Lr, H	0,193
Sn	0,042	Lm,H	0,183
Sкр	0,2	Ks	0,961
Rs,Ohm	1,75	Kr	0,948
Rr,Ohm	1,142	σ	0,0883

Nominal frequency of rotation – Ωn; nominal torque – Mn; scattering coefficient – σ; critical torque – Mcr; stator current – In.; rated slip – Sn; ctirical slip – Scr; stator active resistance – Rs; rotor active resistance – Rr; stator reactive resistance – Xs; rotor reactive resistance – Xr; mutual induction resistance – Xm; stator full induction – Ls; rotor full induction – Lr; rotor and stator mutual induction – Lm; starting torque – Ms; coefficient of stator magnetic connection – Ks; coefficient of stator magnetic connection – Kr.

Methods of getting the nonlinearity curve $Lµ(Iµ)$, the magnetization curve and the dependence of magnetic inductance from magnetization current, experimentally obtained on their basis, for researched engine, are presented in [12,13]

This nonlinearity is described by following polynomial dependence –

$$L(I) = a_0 + a_1 I^2 + a_2 I^4 + a_3 I^6 + a_4 I^8 + a_5 I^{10} -$$

where a_0-a_5 – are approximation coefficients.

a_0=0.12241002; a_1=0.72550627; a_2=0.59; a_3=0.59; a_4= -0.72550627.

For taking into account nonlinearity in the model of vector control system, this dependence should lead to absolute units. In accordance with the methodology [3], excitation current is considered as an additional to base currents ($I_α$, $I_β$) with minus sign. Its is determined from the known formulas from theory of electric drives [14].

Plans models performed in the coordinates α-β-0 [10].

IV. CALCUTIONS FOR REGULATORS

For α-β-0 system, the synthesis of current regulators has a form as followed below. Open-circuit transfer function of current stabilization:

14

$$W_{pa3.к.cma6.I}(p) = W_{PT}(p) \cdot \frac{k_{\Pi}}{T_{\Pi}p + 1} \cdot \frac{1/R_e}{T_{1e}p + 1}; \qquad (1)$$

$$R_{1e} = R_s + k_r^2 R_r; \qquad k_r = \frac{L_m}{L_r}. \qquad (2)$$

A preferred open-circuit transfer function of current stabilization:

$$W_{ж.роз.к.cma6.I}(p) = \frac{1/k_{O.T.}}{a_r T_\mu p(T_\mu p + 1)}; \qquad (3)$$

– where $k_{O.T.}$ - feedback coefficient for the current.

Accepting a_r according to the regulator in setting optimum modular, $T_\mu = T_\Pi$.

Calculating the transfer function of regulator:

$$W_{PT}(p) = \frac{T_{1e}p + 1}{k_{O.T.}2T_\Pi p(1/R_{1e})}; \quad W_{PT}(p) = \frac{T_{1e}p + 1}{T_{ИТ}p}; \qquad (4)$$

$$T_{ИТ} = \frac{k_{O.T.}2T_\Pi k_\Pi}{R_{1e}}; \qquad (5)$$

Writing the transfer function of the open-circuit and closed circuit flow stabilization, for synthesis flow regulator:

$$W_{pa3.к.cma6.II}(p) = W_{P\Pi}(p) \cdot W_{замк.cma6.I}(p) \cdot \frac{k_r L_r}{T_2 p + 1}; \qquad (6)$$

$$W_{замк.cma6.I}(p) = \frac{1/k_{O.T.}}{2T_\Pi p} \cdot W_{P\Pi}(p) = \frac{k_{O.T.1}(T_2 p + 1)}{4T_\Pi p k_{O.\Pi.}L_m}; \qquad (7)$$

Prefered open-circuit transfer function stabilization and flow regulator would be next:

$$W_{6.роз.к.cma6.III}(p) = \frac{1/k_{O.Ш.}}{a_T T_\mu p(T_\mu p + 1)}; \qquad (8)$$

$$W_{PC}(p) = \frac{k_{O.\Pi.}k_{O.T.2}J}{3z_p k_r 2T_\Pi k_{O.C.}} = k_{PC}; \qquad (9)$$

Accepting $a_T = 2$, $a_c = 2$ according to the regulator in setting optimum modular, with decompensated time constant $T_{\mu C} = a_c \cdot T_\Pi$. Time constants $T_{1e} = \sigma \cdot (Ls/R_{1e})$; $T_2 = L_r/R_r$.

Calculation of feedback coefficients is following:

$$k_{O.\Pi.} = \frac{U_{H.3.}}{Y_2}; k_{O.T.1} = \frac{U_{H.3.}}{I_{1a6}}; k_{O.T.2} = \frac{U_{H.3.}}{I_{1\beta6}}; \qquad (10)$$

$$k_{O.C.} = \frac{U_{H.3.}}{\omega_{r.6.}}; k_\Pi = \frac{U_6}{U_{H.3.}}; T_\Pi = 0.001c. \qquad (11)$$

Example calculation of basic values that are listed in the above expressions are given in [10].

Conversion of feedback coefficients are defined by formulas below:

$$\begin{cases} y = \frac{Y}{Y_0}; i_\alpha = \frac{I}{I_6}; n = \frac{\omega_H}{\omega_6}; \Psi = \frac{\Psi_{2H}}{\Psi_6}; \ k` = K \cdot \frac{Y_{ex.6.}}{Y_{eux.6.}}; & (12) \\ k`_{om1} = k_{om} \cdot \frac{I_6}{U_{36}}; k`_{oc} = k_{oc} \ \frac{\omega_6}{U_{36}}; k`_{on} = k_{on} \ \frac{\Psi_6}{U_{36}}; & (13) \end{cases}$$

Calculating coefficients of transfer functions of the block diagram in relating units (active): converter, the function of electric part of motor, speed regulator, Inertial constant of the drive is carried out as follows:

$$\begin{cases} k`_n = k_n \cdot \frac{U_{ex.6}}{U_{eux.6}}; r_{1e} = \frac{R_{1e}}{R_6}; k`_{pc} = \frac{T}{4 \cdot T_n}; \\ T_j = \frac{\omega_6}{M_6} \cdot J_\Sigma; m = \frac{M_c}{M_6}. \end{cases} \qquad (14)$$

For creating a simulation model, we are taking moment of inertia as shown next:

$$J_{мех} = 10 \cdot J_{дв} = 10 \cdot 0.011 = 0.11 \, Н·м;$$

$$J_\Sigma = J_{мех} + J_{дв} = 0.11 + 0.011 = 0.121 \, Н·м$$

According to the expressions, given above, we have following transfer function of regualtors:

$$W_{PT1}(p) = \frac{0.00587p + 1}{0.031p}; W_{PT2}(p) = \frac{0.00587p + 1}{0.021p}; \quad (15)$$

$$W_{P\Pi}(p) = \frac{0.169p + 1}{0.004p}. \qquad (16)$$

V. RESULTS OF MODELING

Results of modeling vector control system in orthogonal axes α-β-0 are given on fig. 4-8.

Fig. 4 – Transient process of AC drive start-up – rotor speed – characteristics with including nonlinearity of magnetization curve (2) and without it (1) (without adjusting moment of inertia of the drive mechanism).

Fig. 4 – Transient process of AC drive start-up – rotor speed – characteristics with including nonlinearity of magnetization curve (2) and without it (1) (with adjusting moment of inertia of the drive mechanism).

Fig. 4 – Transient process of AC drive start-up – stator current – characteristics with including nonlinearity of magnetization curve (2) and without it (1) (without adjusting moment of inertia of the drive mechanism).

Fig. 4 – Transient process of AC drive start-up – stator current – characteristics with including nonlinearity of magnetization curve (2) and without it (1) (with adjusting moment of inertia of the drive mechanism).

Fig. 4 – Transient process of AC drive start-up – flux – characteristics with including nonlinearity of magnetization curve (1) and without it (2).

VI. CONCLUSION

In this analysis of vector control system a research for over viewing a system with orientation by the rotor flux linkage was made. Dynamics characteristics for the constant rotor flux model were build-up. It allows to determine the advantages and disadvantages rather of this method, compared with the scalar operating.

According to the practice (including products of different companies), rather the system in orthogonal coordinates α-β-0 is applied, thus its' calculation of regulators is represented in this paper more detailed. The analysis of the transient processes of the vector system shows, that the influence of the nonlinearity of the magnetization curve of AC drive, results in affecting on the start-up torque of the electric drive, which follows with a little increasing of the time of transient process, and, thus, changing the quality features of the control system – overregulation, regulating time and the time of first maximum.

REFERENCES

[1] Терехов В.М. Системы управления электроприводов: учебник для студ. вузов/ В.М. Терехов, О.И. Осипов; под ред. В.М. Терехова. - 3-е изд., стер. - М.: Издательский центр "Академия", 2008.

[2] Башарин А.В., Новиков В.А., Соколовский Г.Г. Управление электроприводами: Учебное пособие для вузов. - Л.: Энергоиздат., 1982.

[3] Виноградов А.Б. Векторное управление электроприводами переменного тока / ГОУ ВПО "Ивановский государственный энергетический университет имени В.И. Ленина". - 2008.

[4] Рудаков В. В. и др. Асинхронные электроприводы с векторным управлением / В.В. Рудаков, И.М. Столяров, В.А. Дартау. - Л.: Энергоатомиздат, 1987.

[5] Соколовский Г.Г. Электроприводы переменного тока с частотным регулированием: учебник для студ. высш. учебн. зав. Г.Г. Соколовский. - М.: Академия, 2006.

[6] Усольцев А.А. Векторное управление асинхронными двигателями. Учебное пособие по дисциплинам электромеханического цикла. Санкт-Петербург. 2002.

[7] Виноградов, А.Б. Учет потерь в стали, насыщения и повехностного эффекта при моделировании динамических процессов в частотно-регулируемом асинхронном электроприводе / А.Б. Виноградов // Электротехника. – 2005. – №5.

[8] В.Л.Вейц и др. Динамика управляемого электромеханического привода с асинхронными двигателями. – Киев: Наук. думка, 1988.

[9] Куцевалов, В.М. Асинхронные и синхронные машины с массивными роторами / В.М. Куцевалов. – М.: Энергия, 1979.

[10] Козярук А.Е., Рудаков В.В., Народицкий А.Г., Современное и перспективное алгоритмическое обеспчение частотно-регулируемых электроприводов, Санкт-Петербург, 2002.

[11] В.А., Рудаков В.В., Козярук А.Е. и др. Теоретические основы построения частотных электроприводов с векторным управлением / Автоматизированный электропривод – М.: Энергия, 1980.

[12] Калинов А. П. Современные способы определения потерь мощности машин переменного тока / А. П. Калинов // Вісник Східноукраїнського національного університету імені Володимира Даля. – Луганськ: СНУ ім. В. Даля. – 2005. – №1(10) – С. 131-136.

[13] Огарь В. А. Оценка нелинейности индуктивности катушки со сталью энергетическим методом / В. А. Огарь // Вісник КДПУ. – Зб. наук. праць КДПУ. – Кременчук: КДПУ. — 2004. – Вип. 2 (25). – С.78-84.

[14] Т.В. Величко, Д.И. Родькин Теория электропривода. Часть 1. Механика и характеристики двигателей в электроприводе. Кременчуг: КГПИ, 1999.

Capacitive pressure sensor for application in medicine

[1]Michal JURČIŠIN, [2]Pavol CABÚK

[1,2] Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic

[1]michal.jurcisin@tuke.sk, [2]pavol.cabuk@tuke.sk

Abstract — **The first part of paper deals with development of sensor for intra-abdominal pressure (IAP) measuring which does not interfere with body fluid of a patient and thus the keeps sterility of entire system. The second part of paper deals with technical solution of indirect monitoring of IAP by intra-vesicular pressure (IVP) method, which is today regarded as standard for indirect monitoring of IAP. The result of new technique eliminates the influence of a human factor and the possibility to utilize it for urinary bladder tonisation of paraplegic patients.**

Keywords— **intra-abdominal pressure, monitoring, capacitive pressure sensor**

I. INTRODUCTION

Abdominal compartment syndrome (AbCS) is caused by increasing of intra-abdominal pressure (IAP) from tissue tumescence or from cumulating of free liquid in abdominal cavity. AbCS is very hard complication that may impact on many organ systems of patient. When it is not timely diagnosed and it is not cured, it may result in multi-organ collapse and death [1].

The most precision method of the IAP measuring is applying pressure sensor through abdominal wall directly into abdomen cavity. Measuring of pressure directly in abdomen cavity via catheter is a hazardous invasive method, which is not used in clinical practice [2]. Indirect method of IAP measuring is based on Foley catheter inside urinary bladder. This method is based on fact, that urinary bladder works like passive transmitter IAP on internal water filling. Empty urinary bladder is filled by 20-50 ml of sterile physiologic solution via catheter. [2] Pressure in abdomen cavity is transferred across bladder on filled solution and so as on water column of catheter which is connected to water column on a manometer or pressure sensor. [3] Measurement is realized in one hour intervals with filling and empting of all system.

II. CAPACITANCE SENSOR

Capacitive pressure sensors were realized by glass tubes with inner diameter 4-9 mm. On the outer surface of the glass tube was deposited two thin-film electrodes by vapor deposition. As a material for deposition was used copper. The shape of the electrodes is shown in Fig. 1. As dielectric of capacitance sensor was used saline (0.9% NaCl aqueous solution).
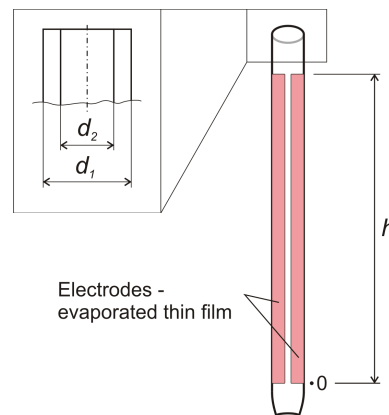
Fig. 1. Capacitive pressure sensor (d_1 – outer tube diameter, d_2 – inner tube diameter, h – height of evaporated electrode

Capacity change of capacitive sensor is measured using a RLC bridge MT4090. LCR meter MOTECH MT 4090 can be connected to computer via RS 232 interface. Using created application can be set any time interval between measurements.

Based on the measurements is possible to directly characterize the capacitive sensor. Analyzed parameters was value of maximal (C_{MAX}) and minimal (C_{MIN}) capacity and the time interval of increase (t_{inc}) and decrease (t_{dec}) of capacity after which is measured capacity in steady state. Capacity C_{MAX} respectively C_{MIN} is capacity measured on the assumption that the level of dielectric is located in the maximum respectively minimum height. The time interval t_{inc} is obtained by forced change of level height and t_{dec} is given by the spontaneous decrease of dielectric level height. On Fig. 2 is shown sample of the test program.
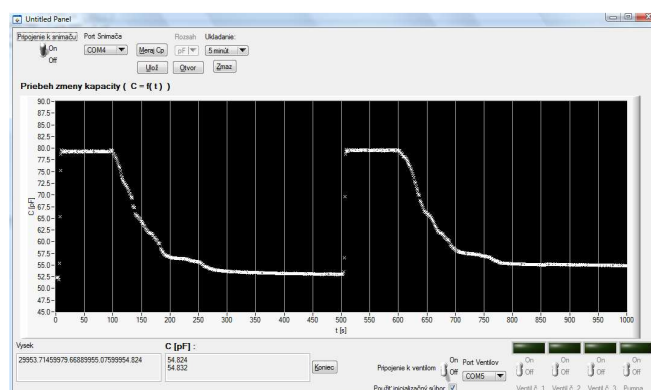
Fig. 2. Measurement course in the created test application

Created application enable shows measured capacity in 1 s interval. The maximum and minimum value of Y axis, on which are plotted the measured values of capacity (pressure) is dynamically changed according to measured values.

In Tab. 1 are listed parameters of used glass tubes. Measured capacity varied depending on the tube diameter in the range of 11,7 pF - 25 pF without dielectric corresponding to the pressure 0 kPa, and 14,3 - 132 pF with a dielectric at height 300 mm corresponding to the pressure 2,94 kPa.

TABLE I
OVERVIEW OF THE TESTED GLASS TUBES PARAMETERS

	Tube No. 1	Tube No. 2
d_1 [mm]	7,3	10,6
d_2 [mm]	3,6	8,65
h [mm]	300	300
C_{MIN} [pF]	~11,7 – ~12,2	25
C_{MIN} [pF]	~14,3 – ~14,9	132
t_{inc} [s]	5	10
t_{dec} [s]	460	65

At the beginning of the measurement was dielectric level of capacitive sensor of pressure on zero point. The dielectric level was then displaced to height h. After capacity stabilization was the water level spontaneously decreased. Fig. 3. and Fig 4. shows the longest permissible period for stable the sensors capacity on a minimum value, which is permitted for developed measuring system.
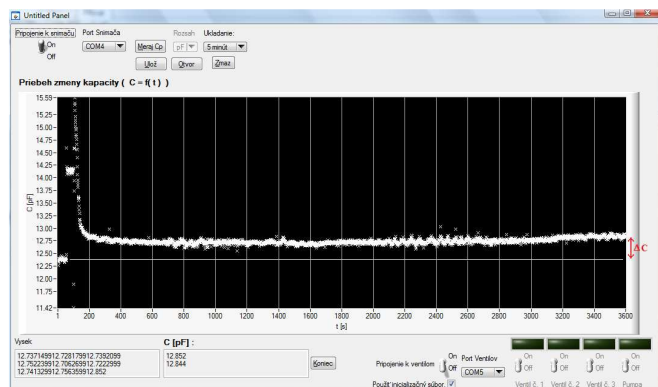
Fig. 3. Capacity course changes of glass tube No. 1
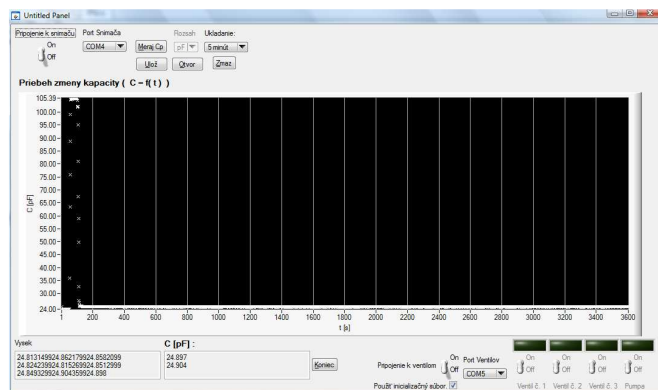
Fig. 4. Capacity course changes of glass tube No. 2

During the time t_{dec} come to stable of capacity at a constant value which did not correspond to the capacity value of C_{MIN} in case of tube No. 1.

In the case of the glass tube No. 1 is capacity range unstable in terms of measurement repeatability. For tubes with these dimensions can be assumed that capacity range is too small for easy allocation a unique value of capacity to dielectric height level.

Capacitance pressure sensor realized by glass tube No. 2 is considered as suitable in terms of all examined parameters. Examined parameters improve with diameters greater than diameter of glass tube No. 2. Using tubes with larger diameters cause increase of internal volume of tube what makes them unsuitable for the application in terms of volume of saline filled into the bladder.

III. DESCRIPTION OF CONCEPT OF MEASURING SYSTEM FOR MEASURING OF INTRA-ABDOMINAL PRESSURE

Main function of developing equipment is automation of measuring process and decreases the time consumption and eliminate human factor as a main factor of uncertainty of measurement. Block scheme of developed measuring system for automation measuring of IAP is shown on Fig. 5. In the next is described projected working cycle. In time interval between measurements are all valves opened whereby are possible to empting of all system and continuously flow of urine to waste. After empting the system is possible to close valve No. 2 for isolation of patient from non-used part of system. This condition decreases possibility of infection. The first step of measurement process is filling tubes between sensor and pump with defined volume of physiologic solution. In the next step are closed valves No. 1 and No. 3 and opened valve No. 2. Bladder is filled with volume (20-50 ml) of physiological solution. After filling start the measurement by opening valve No. 2. Pre-filling tubes around valve No. 3 is ensured continuity of liquid column and the urine interference on sensor is eliminated.

The valves consist of servomotor which can be logically controlled. This solution was realized in order to ensure the sterility of the measuring system.
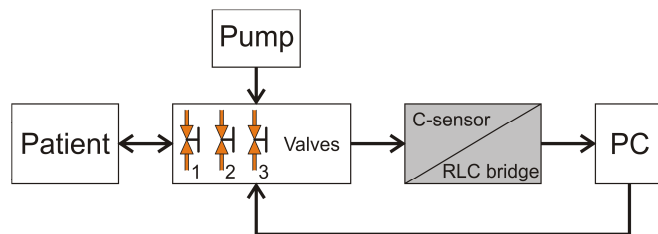
Fig. 5. Principal block diagram of measuring system

IV. CONCLUSION

All the system will simplify measurement process in focus of medical staff and eliminating the human mistakes. In the first step was solved sterility of system, measurement via capacity sensor. Second step was focused to specification of

optimal dimensions of capacitance sensor of pressure. Automated measuring system will be tested in laboratory conditions. (Project VEGA 1/0108/09)

REFERENCES

[1] T. Molčányi, J. Magdo, R. Raši, T. Cestická, J. Michlík, A. Molčányiová: "Možnosti ošetrovania laparostomie při abdominálnom kompartment syndrome" in *Novinky v anasteziológii, intenzívnej medicine a algeziológii*, Prešov, Náuka, 2005, s. 122-127, ISBN 80-89038-37-9.

[2] F. Obeid et al., "Increases in intra-abdominal pressure affect pulmonary compliance", Arch Surg, 1995. 130(5): p. 544-7; discussion 547-8.

[3] Kron IL, Harman PK, Nolan SP: The measurement of intra-abdominal pressure as a criterion for abdominal re-exploration.

14

Classification method of SQL-query as a preliminary stage of parsing

[1]*Pavlo KOSTENKO,* [2]*Mykola GUCHENKO,* [3]*Myroslava IVANOVA,* [4]*Tibor VINCE*

[1]Dept. of Computer and Information Systems, Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine
[2]Dept. of Computer and Information Systems, Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine
[3]Dept. of Computer and Information Systems, Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine
[4]Dept. of Theoretical Electrical Engineering and Electrical Measurement, Technical University of Košice, Slovakia

[1]ppkostenko@gmail.com, [2,3]mig@link.poltava.ua, [4]tibor.vince@tuke.sk

Abstract— **The features of declarative query language SQL is considered. Types of simple and complex SQL-query are specified taking into account aggregate functions, sorting, grouping and joining the tables fields. The classification method of query is developed according to the standard SQL-99.**

Keywords—**DBMS, RDBMS, SQL-query parsing, SQL-query types.**

I. INTRODUCTION

Considerable development and extensive application of computer information systems based on database management systems (DBMS) can be seen in the last years. A special place in such systems is standing by the automated control system workplace (ACSW). Since the release, in 1970-th, the fundamental work [1] by the IBM research laboratory employee Dr. E.F. Kodd about the relational data model, the SQL-queries became the indispensable tool for obtaining information in relational database management system (RDBMS). Structured Query Language (SQL) is a standard communication with the database, which is supported by ANSI. The latest version - SQL-2008, although the standard SQL-200n is in the development. Most databases are firmly adhere to the SQL-99 (SQL3). Despite the accepted standards, commercial database developers widely supplement of the language SQL. Formed unique syntax sets are significantly reducing the flexibility and mobility of the client software.

The incompatibility problem of declarative languages of modern databases can be solved with minor syntax changes at a constant query semantics. Special analyzers programs (parser) implement syntactic changes. Parsers are aimed at finding a limited number of syntactic expressions and its changing with the appropriate standard, adopted by the DBMS. Existing parsers are not without drawbacks, foremost of which is a narrow specific of usage. Structured query language is considered as the industry standard language for RDBMS queries because users need only to specify the required data without going into details on how to access them. Development of flexible parser requires prior detailed analysis of the features of SQL-query because of the declarative properties of SQL.

II. STUDY FILED

In contrast to commercial databases in last years, the usage of universal non-commercial DBMS becomes actual. Open sources of such systems allow to adapt and improve them under special local problem easily. Among the obvious leaders of foreign market are MySQL, PostgreSQL, Sybase ASA. There is only one real "gray cardinal" of domestic space - FireBird. In results of 2007-th [2], FireBird has 13% in distribution of database applications in information systems, it allows to compete with such world commercial brands as Microsoft SQL Server and Oracle..

FireBird is a representative of complex InterBase/FireBird (IB/FB) [3], [4]. This systems don't need many hardware resources, are platform-independent, easy in usage and relatively easy to master. FireBird is free and distributed with the source code, which allows to compile freely database in Linux (and other Unix-systems) using compiler GCC, and to make compiling projects for Windows directly from MS Visual Studio environment..

FireBird database uses standard SQL-2003. According to the type of data processing, SQL is divided into four sublanguages: Data Definition Language (DDL), Data Manipulation Language (DML), Data Control Language (DCL), Data Query Language (DQL). Sublanguages determine the operators list for working with data. DML and DQL have a common operator SELECT. Operator SELECT is designed to return the data collection (select data) from the database, that matching with a given condition. Operator *data query language* is the most widely used and versatile. It complicates the process of parsing and modernization.

The common structure of the query SELECT is:
SELECT
  fields from tables
FROM
  tables for select fields
WHERE
  criterions for selection fields
GROUP BY
  criterions for grouping fields
ORDER BY
  criterions for sorting fields

Described structure creates a large set of options for forming a query, it allows to use of the subqueries for sections SELECT, FROM, WHERE, to construct correlation queries and to combine the tables.

III. SQL-QUERY TYPES

Structure of the SELECT-query allows to distinguish two main types: simple (queries for extract data from unjoined tables) and complex (queries with subqueries, joined table, unions, etc.). Additionally it expands the set of queries in used fields alias, aggregate functions, CASE blocks, grouping and ordering fields. Let's define the basic types of the simple queries:

1. Selecting data from one table.
   select t1.col1, t1.col2, … t1.colN
   from table1 t1
   where condition
2. Selecting data from few tables without joining.
   select
      t1.col1, t1.col2, … t1.colN,
      t2.col1, t2.col2, … t2.colN,
      …
      tM.col1, tM.col2, … tM.colN
   from
      table1 t1, table2 t2, … tableM tM
   where   condition1, condition2, … conditionK
3. Selecting data with using additional operators GROUP BY, ORDER BY.
   select
      t1.col1, t1.col2, … t1.colN,
      t2.col1, t2.col2, … t2.colN,
   from table1 t1, table2 t2
   where condition1, condition2
   group by t1.col1 **order by** t2.col2

Defining types of complex queries requires taking into account possible join options and embedded queries. Subqueries may be included in sections SELECT, FROM, WHERE. In section SELECT, subquery is used to select fields from additional tables and for creating a correlated or uncorrelated query. Subquery in section FROM is used as the internal views. Unlike the subquery in section SELECT, subqueries from section FROM can be performed independently. This subquery doesn't require coordination with external query. As opposed to subqueries from zones SELECT or FROM, subqueries in section WHERE can be coordinated with external queries or uncoordinated. In the first case, they refer to the "current row" of the external query (a typical example of coordinated subquery is "and exists ()").In the second case, subqueries can be executed independently of the external query (a typical example of this is "and column_name in ()").From a logical point of view, they are interchangeable [5].
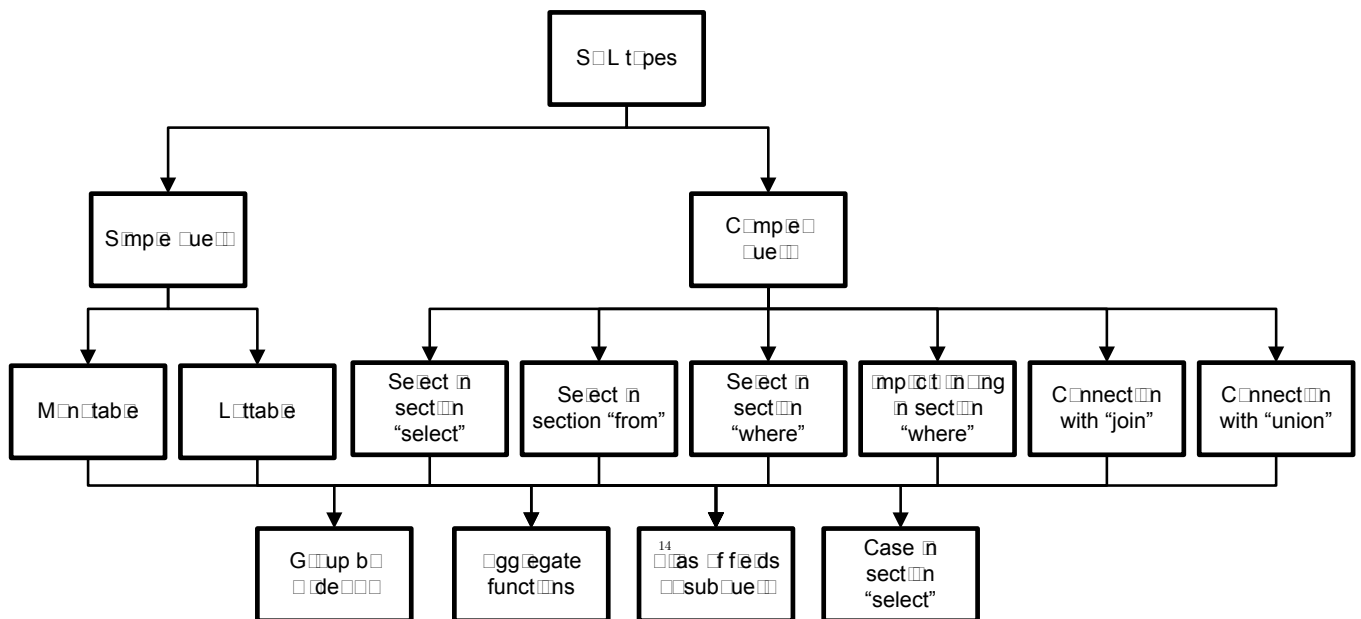
Fig. 1. SQL-query types .

Joining of the queries can be explicit and implicit. Implicit join is realized in the WHERE section. Detection of such variants requires the knowledge of RDBMS metadata, that complicates the parsing additionaly.

   select
      t1.col1, t1.col2, … t1.colN,
      t2.col1, t2.col2, … t2.colN
     from table1 t1, table2 t2
     where t1.PK = t2.FK

Explicit query connection is implemented by the operator JOIN. There are INNER JOIN and OUTER JOIN. In turn, outer join is divided into left (LEFT OUTER JOIN), right (RIGHT OUTER JOIN) and complete (FULL OUTER JOIN). [6]-[9]. It should be noted the opportunity for using the data tables as a set with UNION operator. Full diagram of SQL-query types is shown on Fig. 1. Let us consider the basic types of complex queries with operator JOIN:

1. Query with inner join.
   select
      t1.col1, t1.col2, … t1.colN,

t2.col1, t2.col2, … t2.colN
  from
    table1 t1 **[inner] join** table2 t2 **on** t1.PK = t2.FK
2. Query with left outer join.
  select
    t1.col1, t1.col2, … t1.colN,
    t2.col1, t2.col2, … t2.colN
  from
    table1 t1 **left [outer] join** table2 t2
    on t1.PK = t2.FK
3. Query with right outer join.
  select
    t1.col1, t1.col2, … t1.colN,
    t2.col1, t2.col2, … t2.colN
  from
    table1 t1 **right [outer] join** table2 t2
    on t1.PK = t2.FK
4. Query with full outer join.
  select
    t1.col1, t1.col2, … t1.colN,
    t2.col1, t2.col2, … t2.colN
  from
    table1 t1 **full [outer] join** table2 t2
    on t1.PK = t2.FK

Define the basic types of complex queries with subqueries in sections SELECT, FROM, WHERE.

1. Uncorrelated subquery in section SELECT
  select
    t1.col1, t1.col2, … t1.colN,
    t2.col1, t2.col2, … t2.colN
      (**select** t3.col1 **from** table3 t3
       where t3.col2 = 'GoodName') **as** NewCol
  from table1 t1 **left [outer] join** table2 t2
    on t1.PK = t2.FK
2. Correlated subquery in section SELECT
  select
    t1.col1, t1.col2, … t1.colN
      (**select** t2.col1 **from** table2 t2
       where t2.col2 = 'GoodName'
          and t2.ID = t1.ID) **as** NewCol
  from table1 t1
3. Subquery in section FROM
  select t1.col1, t1.col2, … t1.colN
  from (**select** t2.col1, t2.col2, … t2.colN
      from table2 t2 **where** …) t1
4. Subquery in section WHERE
  select t1.col1, t1.col2, … t1.colN
  from table1 t1
  where t1.col1 = (**select MAX**(t2.col1)
         from table2 t2)

Classification method is developed on the basic of SQL-query types that can detect the local query features and perform it correctly (parsing) analysis.

IV. SQL-QUERY CLASSIFICATION METHOD

The binary grid is used for classification query. The column values in this grid match the unique query types. After query analysis gradual grid filling is performed under checking for the appropriate logical structures. The binary value, that was

converted to decimal form, will provide a unique query. The classification method of SQL-queries is proposed, which consists of the following steps:

1. Validate that the SQL-query belongs to language DQL. (The first query operator is validated for matching to SELECT).

2. Search for second operator SELECT, if it found – move to item 3, item 7 otherwise.

3. Define the query section of additional operator SELECT, return to item 2.

4. Search keyword: INNER, OUTER, LEFT, RIGHT, FULL.

5. Search UNION operator.

6. Search CASE operator in the section SELECT, move to item 10 (complex query)

7. Search keyword INNER, OUTER, LEFT, RIGHT, FULL.

8. Determination of the tables number in the query (appeal to the section FROM), if less than two – go to item 10 (simple query).

9. Define the join conditions between tables in the section WHERE. If the join conditions are not found – query is simple, otherwise – complex.

10. Search availability of the grouping and sorting data.

11. Search aggregate functions and fields aliases.

V. CONCLUSION

The paper identifies the basic types of SQL-query. The classification method of queries is presented, it will allow to develop a "flexible" parser of query language DQL. The analysis of the section WHERE requires further improvement. Question of the usage of complex aggregate functions, such as: EXTRACT (), Between ... and ..., EXISTS (), CAST () remains important and unresolved

REFERENCES

[1] Codd, E.F. (1970). "A Relational Model of Data for Large Shared Data Banks". Communications of the ACM 13 (6): 377–387.

[2] А. Гусев., Какие СУБД прижились в медицине?, CNews.ru: Обзоры и обозрения., http://www.cnews.ru/reviews/free/national2007/articles/systems_databases.shtml

[3] Firebird Quick Start Guide, IBPhoenix Editors, document version 4.2 — covers Firebird 2.5, 2010
http://www.firebirdsql.org/pdfmanual/pdl.pdf

[4] Борри Х. Firebird: руководство разработчика баз данных: Пер. с англ. СПб.: БХВПетсрбург, 2006. 1104 с.: ил

[5] Фаро С., Паскаль Л. Рефакторинг SQL приложений. – Пер. с англ. – СПб: Символ-Плюс, 2009. – 336 с., ил.

[6]   Thomas Neumann: Query Optimization (in Relational Databases). Encyclopedia of Database Systems 2009: 2273-2278

[7]   Фиайли К., SQL: Пер. с англ. – М.: ДМК Пресс, 2003. – 456 с.: ил.

[8]   Kevin Kline, Andrew Zanevsky, Lee Gould, Microsoft T-SQL Performance Tuning Part 3: Query Optimization Strategies, Applications and Database Management Quest Software, Inc. 2005

[9]   Thomas Neumann, Efficient Generation and Execution of DAG-Structured Query Graphs, Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften der Universitat Mannheim, 2005

14

Dynamic phenomena on the external power line conductors

Comparison of the influence of the span size and icing area (May 2011)

[1] *Matúš KATIN,* [2] *Vladimír KRIŠTOF,* [3] *Ľudovít CSÁNYI*

[1]Dept. of Electric power engineering, FEI TU of Košice, Slovak Republic
[2]Dept. of Electric power engineering, FEI TU of Košice, Slovak Republic
[3]Dept. of Electric power engineering, FEI TU of Košice, Slovak Republic

[1]matus.katin@tuke.sk, [2]vladimir.kristof@tuke.sk, [3]ludovit.csanyi@tuke.sk

Abstract— This contribution discuss about the dynamic phenomena on the external power line conductors. The Contribution compares the rate of influence of two factors onto swing up of conductor. First one is influence of the icing area. The second one is the size of span. Both of them can affect the swing amplitude of conductor. The conductor swing simulations were made by using the Cosmos – M program.

Keywords — overhead power line, Cosmos – M, conductor, icing area

I. INTRODUCTION

The main role of the overhead power line is the distribution of the electric power. The electric power is transmitted for long distance from producers to the consumers. We are trying to make this transfer useful, economic and mostly reliable. Considering reliability, the most important thing is influence of climatic condition that affects on overhead power lines. Overhead power lines must be ready to transmit electrical power in any weather. In summer are the overhead power lines loaded mostly with the static load. But in winter to the static load adds also the dynamic loads. Higher dynamic load rises when frost cover is created on the conductors. This frost cover causes additional load of all the elements. The main reason of the swing up of conductor is frost cover of the conductors. This contribution shows the influence of those two factors to swing up amplitude of the conductor.

II. THE CONDUCTOR SWINGING UP MECHANISM

Influence of the suddenly icing fall down occurs to swing up of conductor. After swing up of the conductor will get into balanced position after several strongly damped oscillations. (Fig.1) The impulse for the flaking of icing in most case gives a strong gusting wind or increasing of temperature. In the event, that external power lines conductors are arranged one above other on the spar (spar type" súdok") may result in contact conductors or dangerous approach and subsequently to double-phase short circuit.
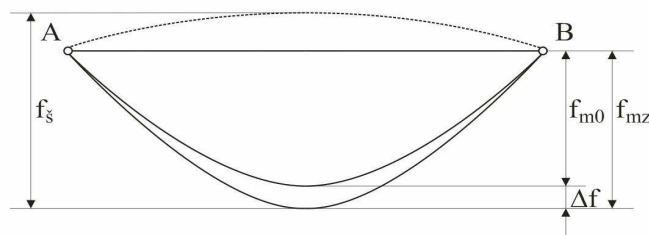
Fig. 1. Swing up of the conductor

III. INFLUENCE OF THE SPAN SIZE

For investigations of span size influence onto swing up of conductor in the simulations was used conductor type AlFe 240/39. We expect only one span in the simulations. The conductor was load with ice cover according to medium icing area.

TABLE I
INFLUENCE THE SPAN SIZE ONTO SWING UP OF CONDUCTOR. MEDIUM ICING AREA

span a [m]	sag in the state 0 f0 [m]	sag at the max. amplitude of oscillation fmak [m]	swing up of the conductor fš [m]
150	2,851	0,669	2,182
200	5,061	2,581	2,481
250	7,916	5,682	2,234
300	11,395	9,372	2,024
400	20,299	18,427	1,872
500	31,782	29,959	1,823
600	45,853	44,062	1,791
800	82,016	80,184	1,832
1000	129,110	127,200	1,910

In the table 1 are given the max. value of swing amplitude. These values reached conductor in swing up simulations. In the picture 2 is illustrated relation between swing up amplitude and span size by the icing fall down.
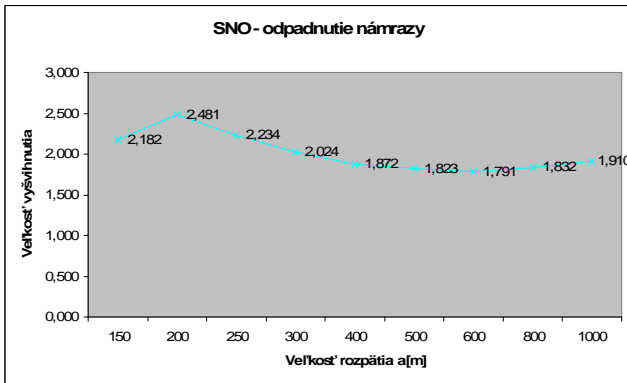
Fig. 2. Graphic dependence of swing up amplitude and size of span. In the swing up simulations was used conductor AlFe 240/39

The following pictures show time behaviour at the swing up of conductor during the icing falling down. These time behaviours were observed in different distances within the span.
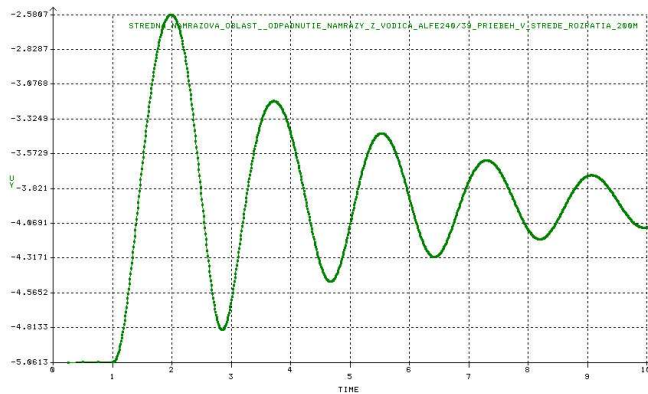
Fig. 3. The time behaviour observed in the middle of 200 m span. In the swing up simulations was used conductor AlFe 240/39
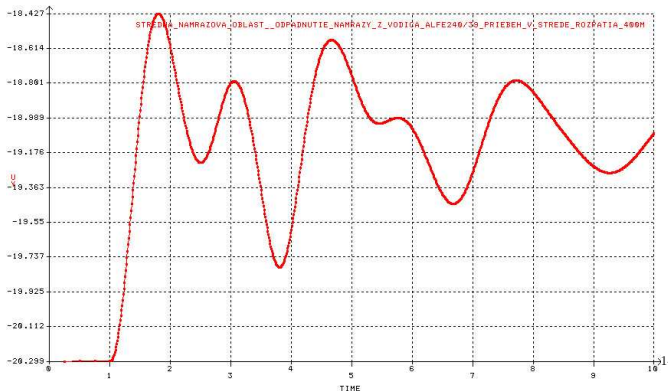
Fig. 3. The time behaviour observed in the middle of 400 m span. In the swing up simulations was used conductor AlFe 240/39

IV. THE INFLUENCE OF THE ICING AREA

For the research how the icing area affected the swing up amplitude in the simulations was used conductor type AlFe 240/39. The maximum allowed stress for this conductor type is 103,807 MPa. According to the standard STN 33 3300 was this value reduced on the 90% level because of the establishment of additional crossings. After this manner was calculated the approximate allowed mechanical stress 93MPa. This conductor has specific strength approximate 273,18 MPa and cross-section 281,66 mm[2]. The value of allowed

mechanical stress 93MPa is representing the condition „0". Condition „0" corresponding to conductor loaded with icing.

TABLE II
WEIGHT AND THICKNESS OF ICING FOR ALFE 240/39 CONDUCTOR TYPE

Icing area	The weight of icing onto measuring wand with 30 mm cross-section	Thickness of icing t [mm]	Weight of icing on 1 m distance [kg/m]	Weight of icing on 1 m distance [kg/m] for conductor type AlFe 240/39 (d=21,75 mm)
L	1	17	0,361+0,0213*d	0,824275
S	2	27,6	0,959+0,0347*d	1,713725
T	3	36,1	1,638+0,0454*d	2,62545
K	All above 3	Counting	1,257*t*(d+t)*10^-3	-

TABLE III
WEIGHT AND THICKNESS OF ICING FOR ALFE 240/39 CONDUCTOR TYPE
ACCORDING TO STANDARD STN EN 50341 – 3 ALOFT 10 M ABOVE THE GROUND

Iging area	The weight of icing onto measuring wand with 30 mm cross-section time to turn 20 years	Reference load with icing IR (N/m) onto unit length of conductor with corss-section d (mm) aloft h = 10 m above the ground	Reference load with icing IR [N/m] onto unit length of conductor AlFe 240/39 (d=21,75 mm) aloft h = 10 m above the ground	Adequate weight of icing [kg/m] onto unit length of conductor AlFe240/39 (d=21,75 mm) aloft h = 10 m above the ground
N0	0,5	1,298+0,1562*d	4,69535	0,47862895
N1	1	3,873+0,2698*d	9,74115	0,992981651
N2	2	10,566+0,4457*d	20,259975	2,065237003
N3	3	18,305+0,5866*d	31,06355	3,166518858
N5	5	35,376+0,8155*d	53,113125	5,414181957
N8	8	63,077+1,0890*d	86,76275	8,844317023
N12	12	102,063+1,3852*d	132,1911	13,47513761
N18	18	162,924+1,7501*d	200,988675	20,4881422
NK	Designate individualy	-	-	-

Table 2 shows weight and thickness of icing responding to icing areas according to old norm STN 33 3300. Table 3 shows reference ice load responding to icing areas according to new European standard STN EN 50341-3. Both of tables contain also computed icing weight for AlFe 240/39 conductor type. These values were used to static and dynamic simulation of swing conductor in aid simulator Cosmos/m. For simulation was used model only of one span and AlFe 240/39 conductor type.

From listed findings we can say follows claims:

> ➢ Maximum amplitude occurs in first swing amplitude in the middle of span. When we have external power lines conductors arranged one above other on the spar, then the distance between any phases is from 3 up to 4 m.
> ➢ In no case of swing up of conductor did not touch the conductors
> ➢ The maximum amplitude of swing up of conductor was bigger when we used in the simulation icing area with smaller reference ice load

TABLE IV
SAGS FOR EACH ICING AREA IN THE MIDDLE OF 300 M SPAN AND THE MAXIMUM SWING UP AMPLITUDE FOR ALFE 240/39 CONDUCTOR TYPE

Icing area	STATE 0 – sag of conductor with icing in corresponding icing area in the middle of span [m] f_{m0}	STATE 2 – sag of conductor without icing in the middle of span after icing fall down [m] f_{m2}	Difference of sags in the state 0 and state 2 [m] Δf	Maximum swing up amplitude of conductor in the middle of span [m] $f_š$
L	7,630	6,532	1,099	2,434
S	11,395	10,215	1,180	2,024
T	15,270	14,228	1,042	1,569
N0	6,170	5,343	0,827	1,811
N1	8,344	7,182	1,162	2,496
N2	12,887	11,756	1,131	1,821
N3	17,578	16,620	0,958	1,384
N5	27,266	26,553	0,713	0,937
N8	42,502	41,967	0,535	0,665
N12	64,386	63,951	0,435	0,544
N18	102,060	101,648	0,412	0,560
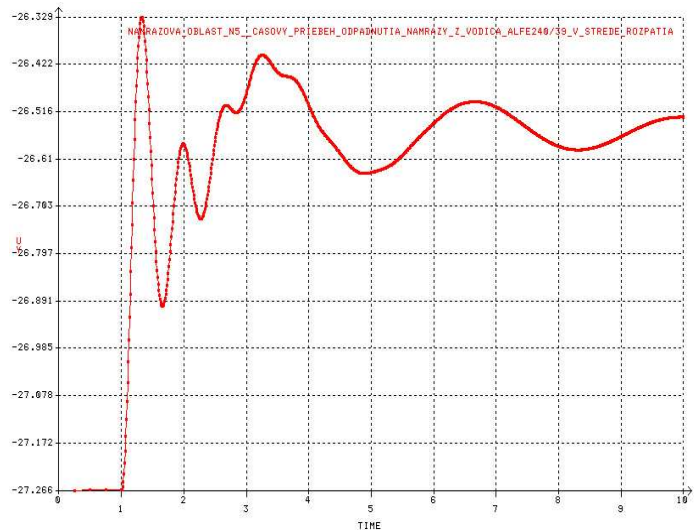
Fig. 3. The time behaviour observed in the middle of 300 m span. In the swing up simulations was used AlFe 240/39 conductor type. And icing area N5

V. CONCLUSION

At the conclusion after evaluation all swing up simulations we can say following results.

The size of span does not have essential influence onto swing up amplitude in the dynamic phenomena of falling down of icing. How we can see in the table 1 as well as in the picture 2. The swing up amplitude of the conductor depends onto span size only slightly. How it can be seen in the case of building new overhead power line we have to give bigger focus on to icing area. In the pictures and graphs can be seen that the swing up amplitude change significantly changing the icing area. The difference of amplitude achieves for single areas the value 2meters.

REFERENCES

[1] R. Hudák, "Riešenie mechaniky kotevného poľa vonkajšieho silového vedenia (VSV) v priestore," Dizertačná práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2010, 130s.
[2] Š. Fecko, "Elektrické siete. Vonkajšie vedenia," SVŠT. Bratislava: Elektrotechnická fakulta, Bratislava, 1978,135s.
[3] L. Varga, - S. Ilenin, "*Elektrické siete*" .Košice,Technická univerzita: 2006, ISBN 978-80-8073-856-3.

Control of AC current of welding machine model using PLC

Peter BARCI

Supervisor of bachelor work: prof. Ing. Dobroslav Kováč, CSc.

Department of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

peter.barci@student.tuke.sk,

Abstract— **The aim of this work is to regulate the power of AC welding machine model, located on the line in the automotive industry. The main objective is the real use of a concretely type of PLC on the welding line in the automotive industry, which represents a large proportion of the overall industry. This work deals with regulating the size of the AC welding current using the PLC (Programmable Logic Controller), a digital computer, which is mainly used in industrial applications in the automation of electromechanical processes. Performance of the circuit consists of a source of welding current and the triac. Regulation in this case provides triac. Control principle lies in the control pulses, which is supplied to the triac control electrode in a set time and precise time intervals. The size of the welding current will therefore depend on the timing of control pulses. The paper describes the performance, and management part of the entire perimeter and the description of generate control signals.**

Keywords - **Automation, triac, PLC, CPU, LAD, FBD, STL SCL**

I. INTRODUCTION

For high technology development and manufacturing industry today is constantly placed greater emphasis on effective management of technological processes in production, whether the car but also each other.

Thus, there is an intense need to improve all manufacturing processes. Substitute for human strength automated production lines, robots and other technological features that optimize production processes. Removing the human factor, the manufacturing processes become more flexible, more efficient, higher quality and ultimately more effective.

A thanks designated benefit of automation is a growing demand management technology. One possibility of management technologies is to use PLC.

II. PERFORMANCE PART

Power semiconductor devices are used almost in all areas of their application used in switching mode. In order to use their main characteristics are essential for our three main states. They are closed, opened and switching status. All these conditions are used in this work.

The aim of this work is to regulate the AC current of welding machine model. For this reason it is used in the work AC inverter, which maintains the frequency of power supply, but changes the effective value of output voltage and current, too. The main element of AC inverter power semiconductor device is a triac.

AC converters

AC converter is an electronic device that changes AC voltage input of certain parameters on the output AC voltage of same frequency but with different effective voltage, and current, too.

The principle of this type of converter is based on the change of "control angle". Description of this converter is explained in Fig.1.

Angle α (control angle) in this case is the point at the pulses are supplied to the control electrode G. At this point the triac gets into the switch on state and on the load is voltage, which corresponds to the actual value of the AC voltage source. Triac and load current begins to flow. In the case of purely resistive load has the same shape as the current tensions. When voltage drops to zero and the current value to the value of retaining current I_H, the triac turns off.

In the second half period voltage source, the triac will behave the same way, because it is symmetrical component. Thanks to these properties, it is clear that delays in the control pulse to the transfer of supply voltage zero will change the effective value of output voltage.
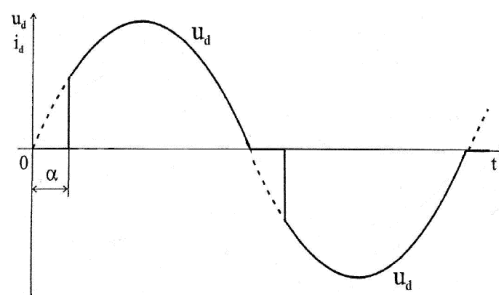
Fig. 1 Course of load voltage

This paper uses the principle described above to resize the welding current. Principle of operation control pulse generator is described in detail in Part III.

III. CONTROL PART

PLC, programmable logic controller is designed and used in industrial applications where it is due to its specific properties broad use. Its main task is to solve complex automation tasks.

This work is carried out using the PLC from Siemens Simatic S7300 CPU313C type (Fig. 2).

Description of the system - a compact CPU, MPI interface, 24 digital inputs/16 digital outputs, 4 analog inputs and 2 analog outputs, an integrated supply 24VDC, 3 fast counters (30kHz), program memory and data: 64kByte.

The main objective of this work is to use the PLC type SIMATIC S7 300 CPU313C with as few external, auxiliary components for generating control pulses for the management of AC welding current.

Fig. 2 PLC Simatic S7-300 CPU313C

There are several kinds of languages for PLC programming. LAD (Ladder Logic) is a graphical programming language using Boolean algebra rules. Logic operations are recorded in that language by linking relay.

FBD (Function Block Diagram), which is also a graphical programming language, linking the blocks that perform standard or custom features. Registration program is similar to patterns of logic and digital circuits. It is very similar language to LAD.

STL (Statement List) is already a text programming language, consists of a sequence of instructions that direct use of the CPU registers the type of PLC. Programming requires a thorough knowledge of the structure of PLC and knows the quantity of commands. Theoretically, the work had to be used, but in this language lacks some necessary features, eg. transfer value WORD to INT, needed to use comparator.

The highest form of programming language for PLC is SCL (Structured Control Language). The SCL is consisting of a sequence of commands. Syntax is similar to the Pascal programming language. This language allows quick and efficient programming without a thorough knowledge of the internal structure of the PLC. This program is an extension of STEP 7.

In this work, all described parts of source code is in SCL, for its simplicity and higher efficiency.

For all programming languages is required STEP 7 program, which allows the creation of projects, hardware configuration, the low-level programming languages, data transfer to PLC and from PLC, and monitoring of selected parameters (Fig. 3).
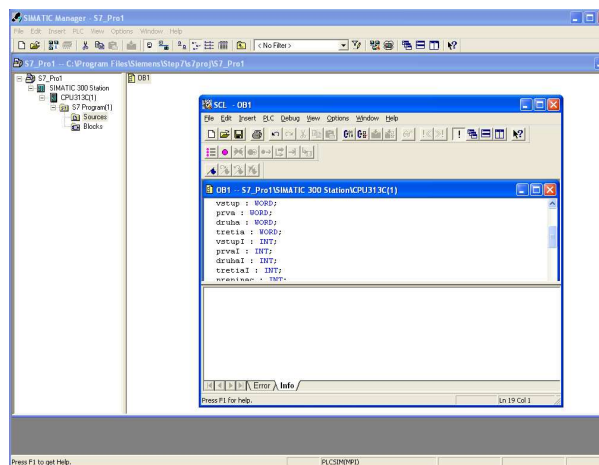
Fig. 3 Program STEP7

IV. DESCRIPTION OF THE SOLUTION

The draft solution describes ways of creating control impulses. In both cases it is important to generate proper impulse sequencing to the instantaneous supply voltage (230V/50Hz).

The principle of solving of the above-described problem lies in connecting the transformer to the one analog input of input / output module and by the PLC control software loading instantaneous (phase) value of power supply.

After stating of zero voltage time the next problem is to generate the firing impulse with the desired control angle.

The first proposed solution consists of using timers by which could generate delayed for correct triac control. In this case, pulses were generated by measuring the time.

The second solution consists in generate pulse by measuring the size of the instantaneous supply voltage.

After studying the parameters of a particular type of PLC we concluded that the first solution is not possible for using PLC type SIMATIC S7 300 CPU313C, because the smallest amount of time for the timer is 10ms. These arguments are based on manuals for SIMATIC S7 300 from Siemens.

For these reasons, it is necessary to generate control pulses by measuring instantaneous power supply.

V. SOLUTION

The measuring of the instantaneous power supply voltage is done via a transformer connected to an analog input of input / output module of PLC. The hardware settings in STEP 7 give us a specific address of an analog input which is connected to the transformer with primary winding 230V/50Hz and secondary winding

14

6V/50Hz. This entry has reserved space in memory inputs I 752.

It is very important to correctly set properties of input module. These settings will find the program STEP7 in the hardware settings of the connected PLC. It is necessary to set the range of analog input values to ± 10 V and suppression of interference frequencies to the value of 400 Hz. This value is set to 50 Hz as default. If we have not exercised this option, so the entire input signal would be suppressed.

Instantaneous voltage is by A/D converter transferred into a binary form (only 12 bit), using the reserved memory space for 16 bit WORD type. Force is thus 12 bits, the remaining 4 bits are the lowest order 0, and they are irrelevant. Conversion time of analog converter is relatively small and it is approximately 1ms. After multiple testing, the practical duration of the transfer time for an average of 20ms was performed 14 to 20 transfers. In this case it is necessary to count with the smallest value, i.e. 14 transfers, representing 7 transfers per half period.

It is therefore possible for a half period to make 7 samples, thus determine the instantaneous values of voltage 7, which represents the ability to create control pulses in seven different times.

Regulating the size of the welding current can thus be at 7 levels. After calculations and even distribution of power are available these reference voltage values: 5.94, 7.67, 8.40. All calculations were performed in Microsoft Office Excel, where you can modify any value, and all the resulting values are converted (Fig. 4).
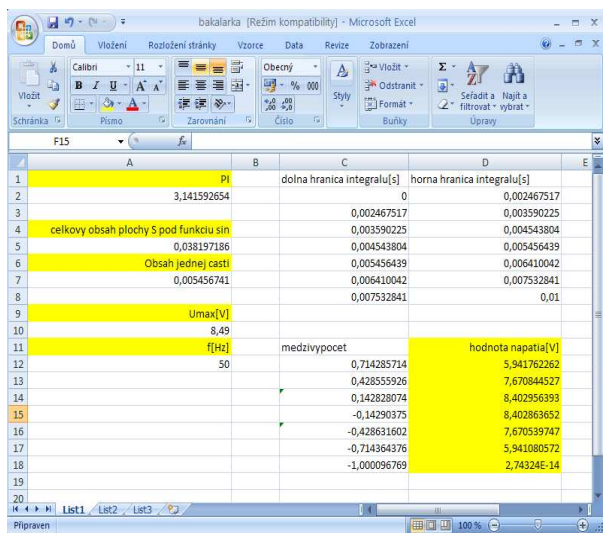
Fig. 4 Calculated limits

It follows that in the moment when the power supply will be located at 5.94 V to get the load 6 / 7 of power. In case of switching triac at a value of 7.67 V the load will get on 5 / 7 of power and at 8.40 V level the load will get on 4 / 7 of power.

Those values are given in the growth shape of supply voltage (first quarter period). In the second quarter of the period the pulses can be generated at the same voltages and load will get on 3/7, 2/7 and 1/7 of power. In the event that the control pulses are generated with the zero control angles the load will get all power.

The principle of the software is based on instantaneous measurements and comparing this value with the calculated values that are stored in a dedicated memory. In the case of a match the measured and set (desired) value is created pulses on digital PLC output whose duration is within the program cycle (ms units).

In the second part of the period the process is the same. Difference is only in storing of calculated constants that are different for negative values.

What is important is fact that the measured values probably never exactly coincide with the calculated, therefore, the program could not compare these values with specific values, but it must to determine in which interval the measured value is immediate.

Another resulting problem is that the calculated values of the voltage are the same even when the voltage grow and fall. It is therefore necessary to establish auxiliary variable that will have the logic value 0, when voltage grow, and logic value 1, when voltage fall. After creating such a variable, we can clearly define whether the measured voltage value is in the first or second half of the half period.

VI. WIRING DIAGRAM
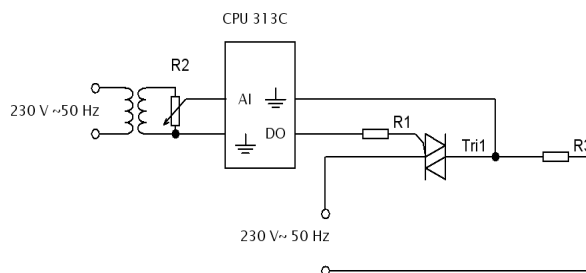
Fig 5 Complete wiring diagram

VII. PARTS OF SOURCE CODE

Retrieve the value of analog input

```
MW10:=PIW752;
vstup := MW10;
```

Saving calculated constants in binary form

```
MW1 :=W#2#01000000_01000000;
prva := MW1;
MW3 :=W#2#01010010_11010000;
druha := MW3;
MW5 :=W#2#01011010_10110000;
tretia := MW5;
```

Convert binary value to INT

```
prvaI :=WORD_TO_INT(prva);
druhaI :=WORD_TO_INT(druha);
tretiaI :=WORD_TO_INT(tretia);
vstupI := WORD_TO_INT(vstup);
```

Interval detection algorithm

```
IF vstupI <= tretiaI THEN
```

```
IF vstupI > druhaI THEN
    Q124.5 :=1;
  ELSE
    Q124.5 :=0;
  END_IF;
   ELSE
    q124.5 :=0;
 END_IF;
```

VIII. CONCLUSION

The aim of this work was to regulate the size of the AC current of welding machine model with the help of PLC. PLC have a very wide application in the automation of industrial processes.

In this case, we were quite limited capabilities of a particular type of PLC. The size of the AC welding current can be controlled in such a case, only after seven stages.

There are several other ways and options to achieve control parameters which can be better.

One of them is the possibility of using external, faster A/D converter, which would be possible to control with much better features. Such an option would be more effective, efficient and with the better quality.

Another option is to use powerful PLC with the timers that have suitable parameters for the practical realization of this work, or a faster input/output module for faster evaluation of analogue signals.

This way is more expensive, therefore using of an external A/D converter is preferable option.

REFERENCES

[1] Fummi, F., Martini, S., Monguzzi, M., Perbellini, G., Poncino, M.: Modeling and Analysis of Heterogeneous Industrial Networks Architectures. *Proceedings of the Design, Automation and Test in Europe Conference and Exhibition Designers' Forum* (DATE'04) 2004

[2] Gu, D., Hu, H.: Neural Predictive Control for a Car-like Mobile Robot. *International Journal of Robotics and Autonomous Systems*, Vol. 39, No. 2-3, May, 2002

[3] Gwynne, P.: Controlling manufacturing via the Internet. url: http://www.spie.org/web/oer/october/oct00/cover1.html created in 2000.

[4] Pelz, G.: Mechatronic Systems. Modelling and Simulation with HDLs, *John Wiley & Sons Ltd, The Atrium*, Southern Gate, Chichester, West Sussex PO19 8SQ, England 2003

[5] Oetter, J.: Power electronics for electrical drives (in Slovak), ALFA Bratislava, 1988

[6] Mohan, N., Undeland, T., M., Robbins, W., P.: Power Electronics, Converters Applications, and Design. *John Wiley*, New York, 1995

Controller simulation model based on a controlled process equivalent

[1]Olena SLAVKO, [2]Mykola GUCHENKO, [3]Myroslava IVANOVA, [4]Ján PERDUĽAK

[1,2,3]Dept. of Computer and Informational Systems, Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine
[4]Dept. of Theoretical Electrical Engineering and Electrical Measurement, Technical University of Košice, Slovakia

slavko.elena@gmail.com, mig@link.poltava.ua, jan.perdulak@tuke.sk

Abstract—The active resonance control algorithm for a stabilizing controller creation is studied. Stabilization of a control system simulation model by a controller based on the active resonance control method in case of low a priori information about a controlled process in real time is demonstrated. Control system simulation model for simulation of mechanical systems stabilization by the active resonance algorithm is created. Advantages and features of the active resonance control method are discussed.

Keywords—Controlled process equivalent, control synthesis, stabilization, simulation.

I. INTRODUCTION

The first task of dynamic systems optimal controlling is a creation of an adequate mathematical model of a controlled object. Structure and parameters of a controlled object mathematical model is partially known or unknown at all for most real systems. In addition, information about external disturbances, which influence on a controlled object and deviate control system's output variable from a defined value, sometimes is insufficient or absent.

So there is a task of a controlled process model creation and its stabilization in real time in case of low a priori information about a controlled process.

As shown in researches [1], [2], human-operator has a unique adaptive ability. Human-operator uses an original approach for a dynamic systems stabilization that differs from technical control systems. Human-operator is able to model an external disturbances signal by its own control signal. This approach is called the active resonance (AR).

In this paper a simulation model of control system stabilization by a controller based on the active resonance control method is created in real time in case of low a priori information about controlled process.

II. PROBLEM OVERVIEW

Control signal can be programmed before the controlling procedure beginning if controlled object dynamics and external disturbance signal are exactly known a priori.

Lack of a priori information about object dynamics and external disturbances can be compensated by current information that is obtained during the controlling process.

In particular control system dynamics prediction can be done on a base of previous dynamics analysis and a controlling signal can be formed on a base of this prediction.

Insufficiency of a priori information about external disturbances, which influence on a controlled object, complicates tasks of identification, simulation, and stabilization of a studied control system.

Using of dual controlling methods assumes that determined or random testing signals are given to the object's input [3], [4], that requires significant time and computational capability and in some cases can not be correctly applied.

Selection of a control algorithm, and controller development are important tasks for controlled object stabilization in case of low a priori information about the controlled process. PID-controllers (Proportional-Integral-Derivative) are known as the most widespread type of controllers for dynamic systems controlling [5].

PID-controllers are mostly applied for controlling of dynamic systems with feedback. But PID-controllers have disadvantages [6] associated with measurement errors of controlled variables, on insufficiency of standard methods of controller parameters tuning and so on.

Procedures of H∞-controllers synthesis within "2 Riccati approach" are known nowadays as a standard [7]. This approach combines a classical theory of automatic controlling and a state space method and allows to set quality and robustness parameters of closed-loop systems in a control process and to change optimal controlling task by suboptimal. The main disadvantage of H∞-controllers is its dependency on a parametric robustness of a controlled object.

Active resonance control approach is a direct method and allows to create an equivalent of unknown external disturbance signal in case of absence of controlled object mathematical model in real time.

Object's stabilization algorithm is the next [2]. Human-operator creates a controlled process model as his own controlling signal that is equal to the influence of unknown external disturbance signal on a controlled object. Human-operator starts to use this model after an external disturbance equivalent creation. Human-operator approximates this created equivalent by some regression function, predicts the

approximated signal and inverts this predicted signal using it as a control signal.

The active resonance control approach will be explained on example of a simulation model of a control system.

III. SIMULATION MODELING

Control system structural scheme (Fig. 1) consists of a controlled object and a controller that forms a controlling signal on a base of the active resonance control algorithm.
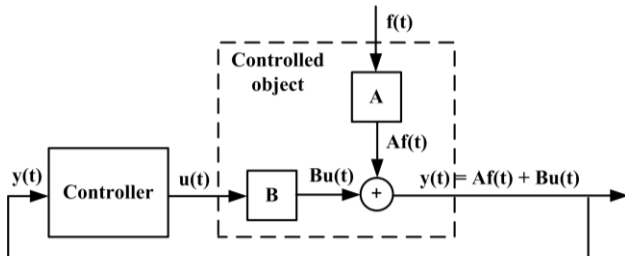
Fig. 1. Control system structural scheme: f(t) – external disturbance signal, u(t) – control signal, y(t) – controlled system output.

Controlled object of this system is linear and described as

$$y\ t\ = A \cdot f\ t\ + B \cdot u\ t\ , \qquad (1)$$

where t – time (s), u(t) – control signal, f(t) – external disturbance, A – amplification coefficient for external disturbance, B – amplification coefficient for a control signal.

External disturbance for this system is a harmonic signal

$$f\ t\ = A_1 \cdot \sin\ \omega \cdot t + \varphi\ + A_0, \qquad (2)$$

where t – time (s), A_1 – amplitude, ω – angle's frequency, φ – phase's shift, A_0 – amplitude's shift.

The active resonance stabilization algorithm is nonlinear and for single iteration of stabilization procedure consists of the next steps:

1) unknown external disturbances equivalent creation;

2) created equivalence approximation by regression function;

3) extrapolation of approximated equivalent of unknown external disturbances for the time interval that is significantly longer than the time interval of the equivalent creation;

4) extrapolated signal inversion.

Matlab / Simulink were chosen for a simulation model creation because of block-hierarchical principle of system models creation with structural-modular programming.

Simulation model of control system with controller based on the active resonance approach is shown in Fig. 2.
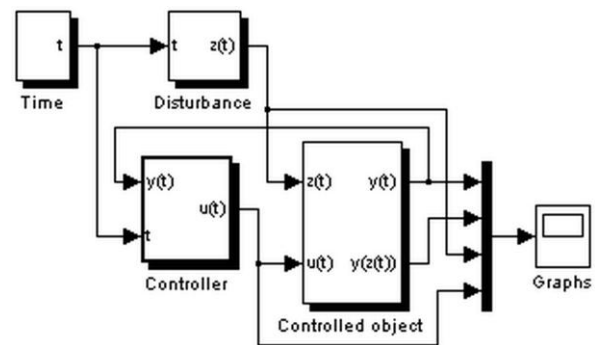
Fig. 2. Simulation model scheme.

Simulation model consists of the next blocks (subsystems): time creation ("Time"), disturbances creation ("Disturbance"), controlled object ("Controlled Object"), controller based on the active resonance control algorithm ("Controller"). Controller consists of three subsystems:

1) unknown external disturbances equivalent creation;

2) external disturbance equivalent approximation;

3) controlling signal forming (extrapolation and inversion of the approximated equivalent).

Example of controlled object stabilization during simulation of a control system (1) is shown in Fig. 3.
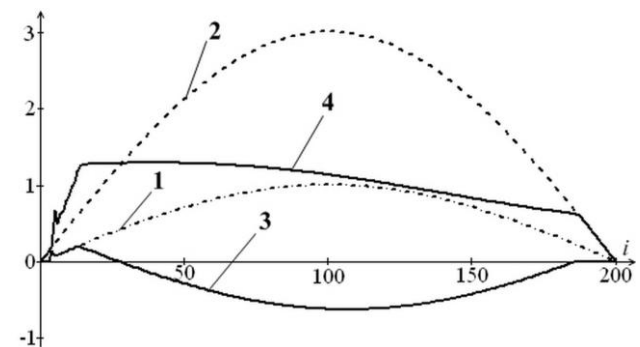
Fig. 3. Example of control system outputs: 1 – external disturbance f(t), 2 – uncontrolled system reaction, 3 – control signal u(t), 4 – controlled system output reaction; i = t/Δt – nondimensional time, Δt = 0,01 (c) – sampling interval, t = 2 (s) – modeling time.

IV. MECHANICAL SYSTEM MODELING

Automatic controllers are modeled in specialized software. The main disadvantage of that software is a theoretical disturbances signals forming without a connection with a real control system. On the other hand, a simulation of a mechanical system under the influence of random external disturbances can be executed in software that is unable to simulate automatic controllers. The requirement to join such software abilities appears for modeling of systems that use the active resonance approach.

Abilities of some simulation systems are well known for temporary information transmission from one system to other by using the OLE- or DDE-technology. But the requirement of the accountability of a simulation real time, model time, its synchronization and unavoidable time delays which are caused by a calculation and data transmission makes an adequate model creation much more complex. One of the active resonance algorithm advantages is its integration with a step-by-step tracking of a system reaction on unknown external disturbances. This allows to create an adequate controlling

signal in a real time.

Controller based on the active resonance control algorithm and a mechanical system simulation can be integrated with modern CAD-systems, avoiding mentioned disadvantages. With this purpose simulation methodology for researches of simulation models of mechanical systems stabilization based on the active resonance control algorithm is developed.

Imitational model creation of a dynamic control system of a mechanical system with the active resonance controller is executed according to the next steps:

1) model and mechanical properties settings;

2) creation of data exchange blocks between controlled object and AR-controller;

3) AR-controller algorithm realization;

4) control signal input to the mechanical system model;

5) model parameters settings and visualization of graphs;

6) settings of calculation accuracy, integration methods, integration step, etc;

7) model simulation, correctness and stability testing.

Dynamic system imitational model creation is based on a mechanical system defined by the following equation:

$$m\frac{d^2x\ t}{dt^2} + \lambda\frac{dx\ t}{dt} + kx\ t\ = f\ t\ +u\ t\ \pm mg, \qquad (3)$$

where $x(t)$ – control system reaction (object's oscillations), $f(t)$ – external disturbance, $u(t)$ – controlling signal, m – controlled object mass, λ – environmental resistance coefficient, k – elastic force coefficient, g – gravity.

The example of a graph of an external disturbance that is unknown for the controlling algorithm and described by a sinusoid in the appropriate script is shown in Fig. 4.
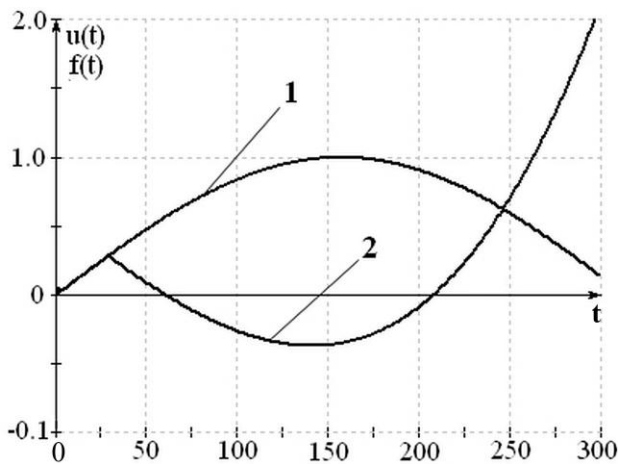
Fig. 4. Example of controlling signal and object stabilization:
1 – external disturbance f(t), 2 – control signal u(t).

Results of the system simulation with and without a control signal are shown in Fig. 5.
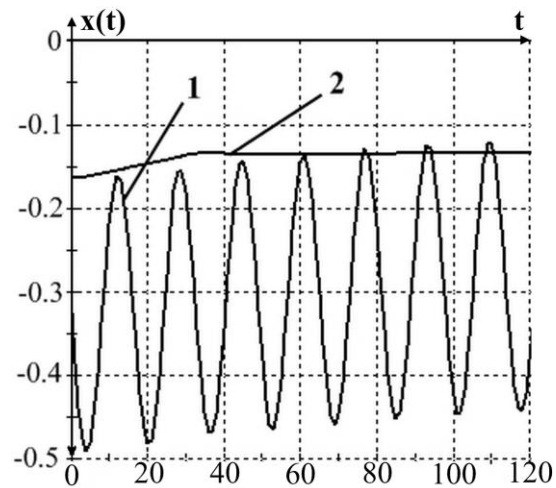
Fig. 5. Example of controlled object oscillations x(t): 1 – uncontrollable, 2 – controllable.

Developed simulation methodology allows to investigate the active resonance control algorithm quality and stability dependences on initial conditions and model and environment parameters for different types of mechanical systems and external disturbances.

Simulation model of control system can be applied for simulation models developing of different types of mechanical systems that can be described by differential or difference equations and external disturbances that can be described by random or determined functions. In addition developed simulation model of control system can be used for researches of simulation models of mechanical systems stabilization by a controller based on the active resonance control algorithm.

V. CONCLUSION

Simulation model of a control system with a controller based on the active resonance control algorithm in real time in case of low a priori information about the controlled process is created in Matlab / Simulink.

Created model allows to realize different methods for approximation and extrapolation of different external disturbances and to investigate which methods are optimal for control system stabilization by the AR control algorithm.

Created model is useful for defining of possible parameters of linear control object and harmonic disturbances and for investigation of nonlinear objects stabilization by the active resonance control approach.

The active resonance control algorithm allows to create adequate controlling signal when connection with a control object is lost for some time interval and possibility to receive system output signal is absent or this possibility is present only in some time intervals or moments.

Results of this paper are using in developing of controllers based on the active resonance control approach. Obtained results will be applied for the further investigation and simplification of a mechanical systems creation of different types of systems with the AR controller, for quantitative and qualitative estimations of the active resonance controller for control systems with random external disturbances.

REFERENCES

[1] M.I. Guchenko, "Active resonance as a principle of operator reaction on the input signal in a compensatory tracking system (in Ukrainian)," *"System technology" Interdepartmental Scientific and Technical Journal,* no. 4'(24), pp. 30–38, 2001.

[2] M.I. Guchenko, "Active resonance stabilizing algorithm (in Ukrainian)," *Bulletin "New technologies" of economics and new technologies institute of J.I. Kravchenko*, no. 1(2), pp. 57−61, 2003.

[3] L. Ljung, *System identification. Theory for the user. Second edition*. New Jersey, USA: Prentice Hall PTR, 1999.

[4] P. Eykhoff, *System Identification: Parameter and State Estimation*. New York, USA: John Wiley and Sons, 1974.

[5] K. Astrom, T. Huggland, *PID Controllers: Theory, Design and Tuning. 2nd edition.* North Carolina, USA: Instrument Society of America, Research Triangle Park, 1995.

[6] H. Lutz, W. Wendt, *Taschenbuch der Regelungstechnik*. Auflage Frankfurt am Main: Harri Deutsch Verlag, 2005.

[7] N.D. Egupov, K.A. Pupkov, Methods of Classical and modern theory of automatic controlling. Controllers Synthesis of Automatic Control Systems (in Russian). 2nd edition. Moscow, Russia: MGTU, 2004.

14

Converter systems simulation in Orcad/PSpice and Matlab/Simulink programs

[1]*Erik EÖTVÖS,* [2]*Milan LACKO*

Department of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]erik.eotvos@tuke.sk, [2]milan.lacko@tuke.sk

Abstract — **The paper describes possibilities of the power circuits' simulation in programs Orcad/PSpice and Matlab/Simulink with SimPowerSystems Toolbox (SPS). Next the advantages and disadvantages of both programs and their cross connection are described. The very effective simulation tool for complex circuits with their control stages is obtained by cross connection of Matlab and PSpice, for example for simulation of the switching power supplies.**

Keywords—**Matlab/Simulink, Orcad/PSpice, PFC – power factor corrector, SLPS–SimuLink PSpice, SMPS - Switched-mode power supply**

I. INTRODUCTION

Nowadays many simulation programs exist for electric circuit simulation (Spice), where from the simplest circuits simulations to the very complex circuits simulation can be done. Assuredly the advantage of these simulations is their high accurate because of very precisely described models of components used in simulation. These models are still filling up and so the simulation with the newest components is no problem.

Simulation program can be use for simulation of the analog control circuit in the switching power supply and for individual power circuit too. Now with power increasing of DSC microcontroller (digital signal controller) is more often the digital control circuits used for the control of converter system. Therefore the applying of digital control circuit in simulation is very suitable. But the many problems can arise from simulation of complex control circuits and especially from simulation of different control structures in simulation program. The problem of simulation convergence occurs very often due to the very complex simulation scheme. A modification of control circuit can be a problem also. The impossibility of logical control circuit simulation in discrete area is a big disadvantage of "spice" simulation program [1].

Matlab/Simulink program with SimPower Systems toolbox (SPS) provides a wide range of possibilities by simulation of electrical circuits. A huge advantage compared to "spice" simulation is ability to create a complex control circuit with common blocks in Simulink. As a consequence the design and modification of simulation circuit is very simplified. On the other hand the SPS toolbox uses very simplified component model which can run under special conditions in real-time too. Furthermore, there is not possible to analyze the waveforms of switching processes or their power losses in SPS models. Therefore the simulation of SPS models is not so accurate than "spice" simulation [4].

By comparison of both simulation tools we can find that every system provides an advantage that is a problem for another system. If this two different simulation tools could be linked together and from both system the only positive properties could be used, so the very strong, powerful simulation tool will be obtained for simulation of the complicated circuit, for example switching power supply.

This possibility is provided by simulation program OrCad/PSpice by Cadence® and Matlab/Simulink®. Their linking together is assured by Matlab block SLSP (SimuLink/PSpice). In the next part the comparison of simulation in OrCad/PSpice environment and Matlab/Simulink/SPS one is described. The main features are presented on the buck converter.

II. BUCK CONVERTER SIMULATION

A. Orcad/PSpice Simulation

The buck converter simulation scheme is shown in Fig.1.
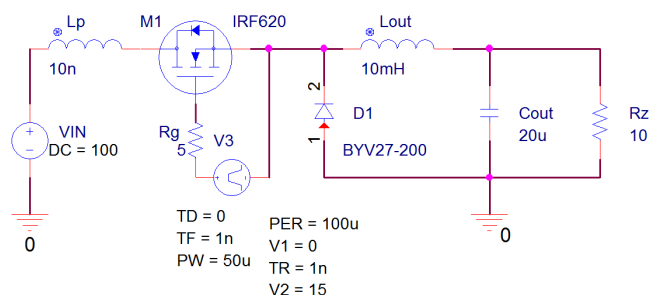
Fig.1 The Buck converter simulation scheme in Orcad/PSpice

The converter is fed by a DC power supply with voltage 100 volts. As a switch the MOSFET transistor is used which is switching with frequency 10 kHz and duty cycle is 50 per cent. In front of switch the inductance L_P was connected that illustrated a parasitic inductance of wire. Next the LC filter is created by inductance L_{OUT} and capacitor C_{OUT} and of course a freewheeling diode. The ohmic load is connected to the converter output. In Fig. 2 the voltage waveform and the current waveform of the buck converter are shown. By the transistor turn-on the current peak is shown on the current waveform, which is caused by long recovery time of freewheeling diode. The switching transients are highlighted in the Fig.2 [2].

On the voltage waveform, the small overvoltage peak is

shown due to the parasitic inductance of the wire, when the switch is turned off.

Simulation results are transformed to the Matlab environment because of unification of the result waveform.
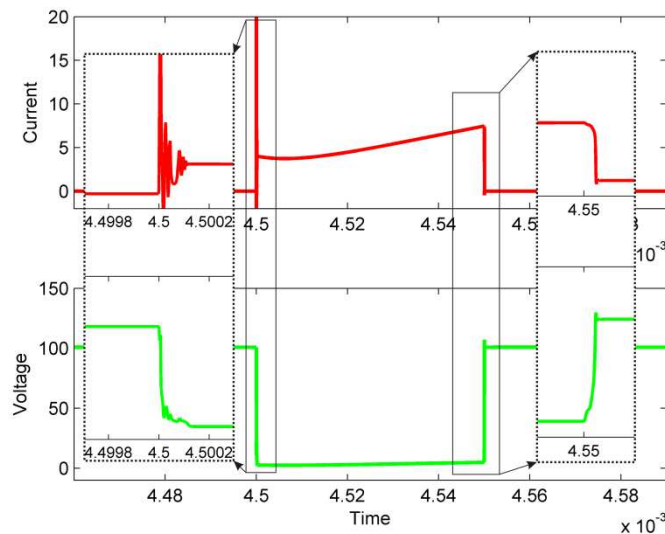
Fig. 2 Buck converter voltage waveform and current waveforms on power switch in Orcad/PSpice

B. Matlab/Simulink Simulation

Similarly as by Orcad/PSpice simulation the buck converter simulation was done in Matlab/Simulink/SPS toolbox. The created simulation scheme is shown in Fig. 3.
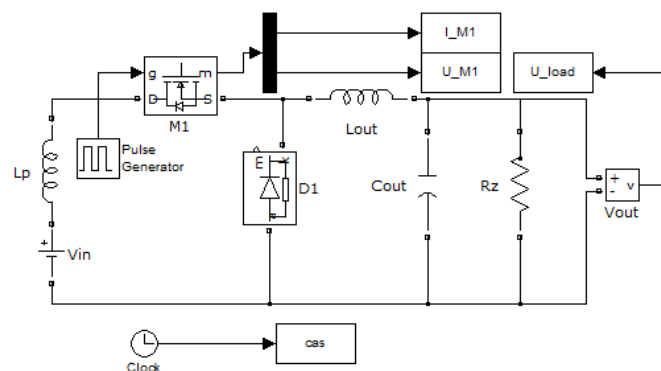
Fig. 3 Simulation scheme of the buck converter in Matlab/Simulink SimPower Systems

Principle of connection is the same as in the previous case. The basic difference is in used semiconductor elements. SimPower Systems toolbox provides only semiconductor elements, which have almost ideal behavior (characteristic). Therefore there is the relatively high inaccuracy of simulation and especially inaccuracy of switching processes.

Obtained simulation waveforms are shown in Fig. 4 in Matlab/Simulink environment. In Fig. 4 we can see the waveforms are practically ideal, what is demonstrated by detail of power switch turning-on and turning-off process. This simulation demonstrates that the Matlab/Simulink environment is not suitable for power circuit simulation and especially if we are interested in fast transient response what just turning process of power switch absolutely is.

But if we want to analyze theoretical behavior of a circuit we do not need to consider elementary actions in analyzed

circuit, thus Matlab/Simulink is a perfect simulation tool. If we need to design a control circuit so the creation of regulation structure would no problem.
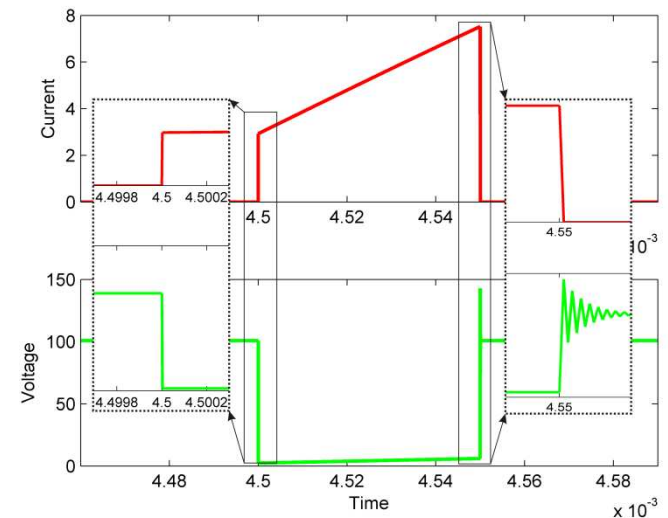
Fig. 4 Buck converter voltage waveform and current waveforms on power switch in Matlab/Simulink

III. CROSS CONNECTION OF THE SIMULATING PROGRAM ORCAD/PSPICE AND MATLAB/SIMULINK

In previous part of paper the advantages of both different simulating environments were described. The simulation of simply buck converter was done. The converter was fed by constant voltage and power switch was switched with constant duty cycle and constant frequency. The load of the buck converter was constant, too.

If the request to circuit simulation with variable condition is exist, so we have to implement the suitable controller to simulation circuit. If we would simulate this circuit in OrCad/PSpice a problem with design and simulation of controller circuit occurs. On the other hand, by using Matlab/Simulink no problem would occur by creating control circuit, although this one could be more complicated. The problem would be just power circuit simulation because of high inaccuracy of elements' models compared to PSpice simulation.

The advantages of both programs will be achieved when the both simulation programs will be cross connected by SLPS block in Matlab/Simulink. So the high precision (accuracy) and simply possibility to create regulating circuit will be achieved.

In buck converter case, where output voltage is regulated parameter, we need to create a controller which changes a duty cycle of the power switch according to desired output voltage. However, we know control circuit, where the sensing of several circuit parameters is needed to control the one parameter. A good example is PFC preregulator, where to control of output voltage the input and output voltage and input current is needed to sense. The regulating structure of PFC preregulator is shown in Fig. 5. The corrected output voltage value is fed into PI regulator. The multiplication block is fed by PI regulator output and by output current value or if you like inductor current. Result of multiplication is fed into comparing block and here it is compared with actual value of output voltage. The output logical value controls the RESET

input of SET/RESET block and defines the switch turn-off point. Into SET/RESET block the clock signal to SET input is fed, which defines switching frequency of the power switch [3].
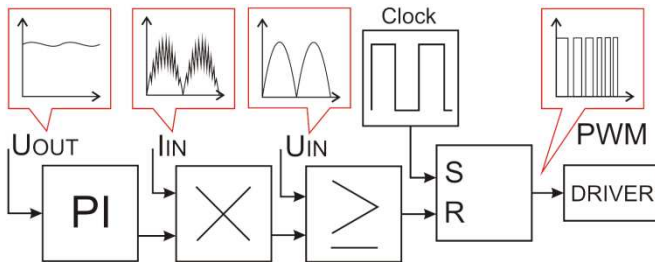
Fig. 5 Regulating structure of the PFC preregulator

The PFC preregulator circuit was created in Orcad/Pspice program (Fig. 6). The diode rectifier bridge is fed by input voltage 230 volts 50 Hertz through input LC filter. Behind the rectifier bridge the small filtering capacitor C_{IN} is placed to filter disturbances caused by power switch. Next the filtering inductor L_{IN} connected. In this inductor the energy is accumulated during turn-on state of the switch Q_1. Rectifier diode D_5 follows the filtering. The output capacitor C_6 is used to smoothing the PFC output voltage. A load with controlled voltage is connected to the converter output. We can control every power supply used in scheme by SLPS block and Matlab/Simulink. The output parameters like current, voltage or power of every single Pspice elements could be transferred to Matlab/Simulink.
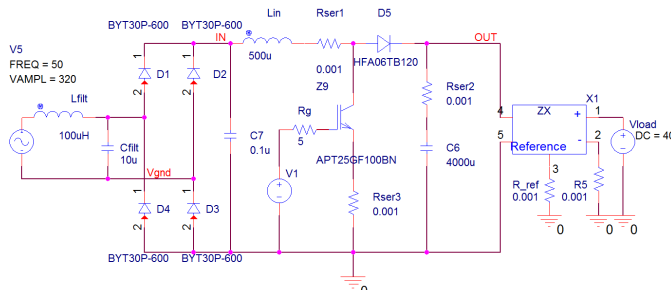
Fig. 6 Simulation scheme of the bridge rectifier with PFC preregulator

After creating the power circuit scheme in Orcad/Pspice, the control circuit was arranged in Matlab/Simulink. Across connection of both programs was performed by SLPS block (see **Fig. 7**).
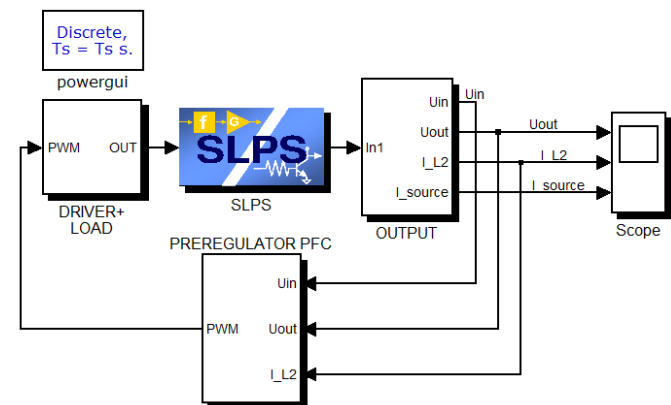
Fig. 7 Cross connection of regulator and block SLPS in Matlab/PSpice

In SLPS block it is possible to select controlled power supply and requested output signals. Then the whole

connection was simulated and the simulation results were obtained. In Fig. 8 the characteristic waveforms are shown. In top part of figure the filtering inductor current waveform is shown. The average value of this current has almost sinusoidal shape of bridge rectifier. Below inductor current the input rectified voltage and input current are shown. On the bottom figure the output voltage is displayed. The ripple of output voltage is ±5 Volts what represents ripple 2.5 per cent of output voltage 400 Volts.
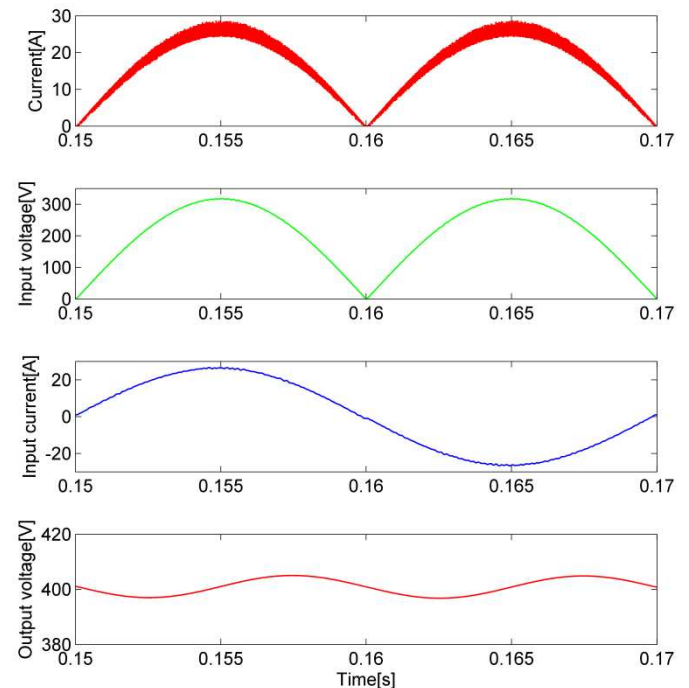
Fig. 8 Waveforms of the PFC pre regulator, filter inductance current, grid voltage, input current and output regulated voltage

The shape of load current form is sinusoidal (see Fig. 8). Furthermore, we can use next tools in Matlab/Simulink environment for better analyses of modeled system (e.g.FFT analysis).
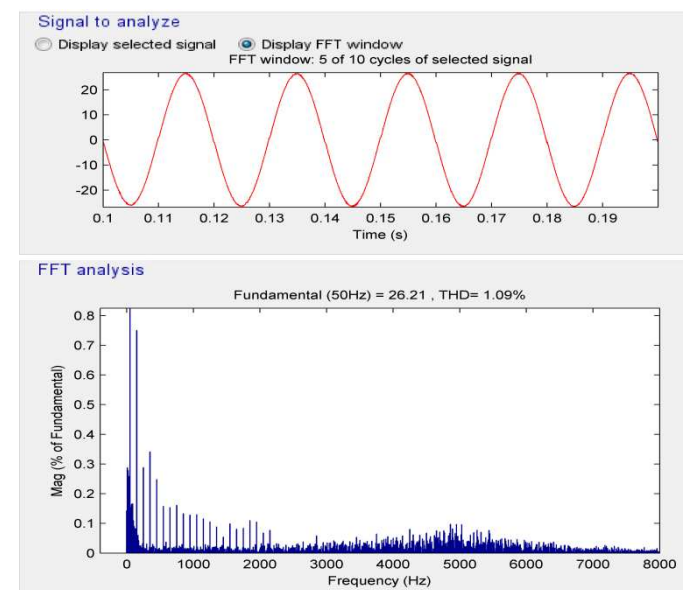
Fig. 9 FFT analysis of PFC preregulator input current

The output current was analyzed by FFT analysis in time from 0.1 to 0.2 second. On the bottom part of figure Fig. 9 the

particulate harmonic components are shown. The fundamental harmonic (50 Hertz) is 100 per cent (out of diagram range). The upper odd harmonics follow (3th - 0.74%, 5th - 0.30%, 7th - 0.34%). The upper harmonics are appeared in switching frequency area, too. A total harmonic distortion is 1.09 per cent what introduces a perfect value.

IV. CONCLUSION

The programs Matlab/Simulink and Orcad/PSpice are suitable for different types of simulation. It was compared by converter simulations in both programs. Orcad/PSpice is better for detailed analyses of switching phenomena without complicated control circuits. But if we need to simulate a detail behavior of power stage of the circuit together with complicated regulating circuit, the cross connection of that program is very desirable. So very strong tool for electrical circuit analyze is obtained.

Also the Matlab/Simulink/SPS provides many auxiliary functions, those can be use by next data processing and simulation analysis.

REFERENCES

[1] P. Tobin, "PSpice for Circuit Theory and Electronic Devices", ISBN: 1598291564, 2007
[2] N. Mohan, T. M. Undeland, W. P. Robbins, "Power Electronics", ISBN: 0-471-58408-8, 1995, ch. 7
[3] M. H. Rashid, "Power Electronics Handbook", ISBN: 0125816502, 2001, ch. 19
[4] Opal-RT Technologies, Inc., *RT-LAB Electrical Applications*, 2007
[5] The MathWorks, Inc., *Power System Blockset*, 2000

14

Data association methods for target tracking by radar sensor network

Daniel Urdzík

Dept. of Electronics and Multimedia Communications, Technical University of Košice, Park Komenského 13,041 20 Košice, Slovak Republic

daniel.urdzik@tuke.sk

Abstract. *The objective of this paper is to provide a survey of basic methods of multiple target tracking by radar sensor network. The tracking of multiple targets by radar sensor network is connected with problems if targets tracks are crossing or if the targets are close spaced. In order to solve the problem the data association methods can be used with an advantage. The specific attention is paid to the underlying principle of basic data association methods used in multi target scenarios.*

Keywords

data association, data fusion, multiple target tracking, nearest neighbor, probabilistic data association, multiple hypotheses tracking

1. Introduction

The modern surveillance systems used for safety, military or control applications often employ the radar sensor networks. The radar sensor networks unlike one radar system can cover larger area of surveillance, and can improve detectability and tracking of the targets. The example of use of radar sensor networks is the tracking of multiple moving people.

In typical tracking procedure, the system receives a set of measurements such as positions of the target at regular time intervals, and forms the track of the target. A problem[14] arises when multiple targets are to be tracked by radar sensor network. Each sensor of the sensor network creates a set of measurements that belongs to the targets. In such situation it is necessary to associate the measurements produced by different sensors to the right targets. Another problem is to distinguish the targets when they are close spaced or if their tracks are crossing. To solve these problems the tracking procedure is need to extend by *data association*.

Data association is an important since it determines the efficiency of the tracking algorithms. If the correct association are made it can improve the tracking accuracy and observability of the targets. Association between a track and the measurements is done whenever their distance is below a defined threshold [1]. In data association algorithms the distance is Mhalanobis distance. The associated measures are then used in tracking to update the predicted states of the targets.

Data association methods could be classified into three groups. They differ in the way when and how the data association is performed e.g. before tracking or after tracking. First the target-oriented approach (plot-to-plot) assumes that the data (coordinates of the targets) are the set of data that have to be associated first, before the tracking process starts [2]. If the correct coordinates of the targets have been associated to the target, then the tracking procedure is applied to track multiple targets. The track-oriented approach (plot-to-track) is the next group of data association [3]. In this case the multiple tracks of multiple targets exist. The incoming measurements are associated to the targets and then these measurements are used to update the existing tracks. The measurement-oriented approach (track-to-track) is another possible solution of data association. In this case it is assumed that multiple tracks of multiple targets exist. Furthermore it is assumed that more than one radar system is present. Each of the radar system creates a different set of tracks and the goal of the track-to-track data association is to associate the correct track from the different radar systems to the correct target. In the literature this approach is described as the *data fusion* process [4]. However track-to-track fusion is beyond the scope of this work and will not be described.

This paper is organized as follows. In section 2, several basic data association methods are described with their modifications as the multi target tracking solution. Section 3 contain the conclusion and future work.

2. Basic methods of data association

Every target dynamics state evolves according to the certain model and that model is used in the tracking filter. Here the Kalman filter will be used as the tracking filter. Kalman filtering consists of two steps, the step of prediction and the step of the update. Prediction step uses the previous state estimation to produce *a priori* state

estimate at current time instant [5]. A *priori* state estimation does not include information from measurements produced by sensors. The update step the state estimation is improved by incoming measurements and hence the *a posteriori* estimation is obtained. The measurement at the time instant *k* is described as:

$$z(k) = H(k)x(k) + v(k), \qquad (1)$$

where the value $H(k)$ represents observation model and v_k is the observation noise assumed to be Gaussian. The prediction step is described by following equations:

$$\hat{x}(k|k-1) = F(k)\hat{x}(k-1|k-1) + B(k)u(k), \quad (2)$$

$$P(k|k-1) = F(k)P(k-1|k-1)F(k)^T + Q(k). \quad (3)$$

The $\hat{x}(k|k-1)$ and $P(k|k-1)$ represents the state prediction and the covariance matrix of the prediction error respectively. The plot-to-track data association utilizes the prediction stage information which is used for association of the measurements to the track [6]. The a priori state estimate and its covariance matrix define the validation region for each target track. In literature this is described as *gating* [7]. The measurements which fall into the gate (validation region) are the input data for tracking. The measurements outside the gate are neglected and are not associated with the track and hence are not used for the track estimation.

2.1 Nearest neighbor data association and its modifications

The nearest neighbor (NN) algorithm associates the measurement which is closest to the predicted state (e.g. position) [8]. The association is based on the Euclidean distance between the current measurement and predicted position (2). Gating of the NN algorithm is expressed as:

$$d_{ij}^2 = \tilde{y}^T(k)S^{-1}(k)\tilde{y} < G. \qquad (4)$$

The value *y* is the measurement residual (innovation). The $S^{-1}(k)$ is the inverse innovation covariance matrix. It describes the influence measurement uncertainty of the noise of the sensor system. The *G* is the threshold or the gate. If the distance of measurement from the predicted state is greater than *G,* it is neglected.

If the measurements from multiple targets fall within the same gate the NN algorithm cannot provide reliable associations. In the Fig. 1 such situation is shown where P1, P2 and P3 are predicted target positions of the targets respectively and O1, O2, and O3 represents measurements from that targets and circles represents range gates. In such conflict situations the NN algorithm gives false results of association.
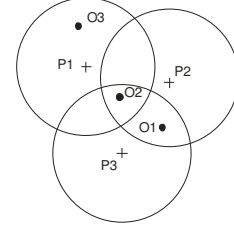
Fig. 1. Conflict situation with measurements within the same gate

The situation where multiple measurements from multiple targets are have to be associated with multiple tracks can be handled by modification of NN algorithm known as global nearest neighbor algorithm (GNN) [9]. Let us suppose that there are *i* tracks and *j* observations. The GNN solves the problem of measurement association through the assignment matrix with dimension of *i* rows and *j* columns. The elements of this matrix will have the following values:

$$d_{G_{ij}}^2 = d_{ij}^2 + \ln(|S_{ij}|), \qquad (5)$$

where d_{ij}^2 can be computed by (4) and $\ln(|S_{ij}|)$ is the logarithm of the determinant of the innovation covariance matrix. The equation (5) is known as the generalized statistical distance. The measurement that satisfies the gating (4) are used to form the assignment matrix. The desired solution of association is the one that minimizes the summed total distance which means that only one measurement from the assignment matrix is valid and all other neglected. In simple cases it can be computed by enumeration but in more complicated situations than depicted in the Fig. 1 the Munkres algorithm can be used as the optimal solution [9].

Other widely used modifications of NN are the strongest nearest neighbor (SNN) [16], and probabilistic SNN [17] which select the measurement in the gate based on the amplitude of the signal. This however assumes that the signal intensity is known and that it is used for the future state estimation.

2.2 Probabilistic data association methods

[14] The different approach of data association is the probabilistic approach known as probability data association (PDA) [10]. Let us suppose that one track exist and Kalman filter is used as the tracking filter. Further we can denote this track as *j* and a set of measurements are to be associated with this track. The PDA was designed for such case. The process of association starts like in the case of NN data association by forming a gate (4), which creates a region where *m* validated measurements are present. Then for each of the validated measurements two hypotheses can be assumed. The hypothesis H_i represents the fact that the i-th measurement z_i has originated from the target. The hypothesis H_0 is the case where no measurement has originated from the target. Hence for the track we can form

m+1 hypotheses. The probability of occurrence of the hypothesis H_0 is computed as:

$$p_{i0}^{'} = \beta^m(1 - P_D P_G).$$ (6)

For the i-th measurement the probability of occurrence of the hypothesis H_i is computed as:

$$p_{ij}^{'} = \frac{\beta^{m-1} P_D e^{\frac{-d_{ij}^2}{2}}}{(2\pi)^{m/2} \sqrt{|S_{ij}|}}.$$ (7)

The β is the probability density function of the occurrence of the false measurement in the validation gate, P_G is the probability that the correct target is present in the gate and the P_D is the probability of the target detection. The values S_{ij} and d_{ij}^2 are the same as in (4) and (5). The probabilities that assign observation i to the track j are computed as follows:

$$p_{ij} = \frac{p_{ij}^{'}}{\sum_{l=0}^{m} p_{il}^{'}}.$$ (8)

The next step after the probabilities of association are computed is the combination of hypotheses for the Kalman filter update equation. The combined innovation is the weighted sum expressed as:

$$\tilde{z}_j = \sum_{i=1}^{m} p_{ij} \tilde{y}_{ij},$$ (9)

where

$$\tilde{y}_{ij} = z_i(k) - H \hat{x}_j(k|k-1).$$ (10)

Then the combined innovation is used in the Kalman filter update step:

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K(k)\hat{z}_j(k).$$ (11)

The rest of the PDA algorithm lies in the computation of the modified covariance P [11], [8].

The PDA algorithm showed significant disadvantages[14] such as in situations where multiple target tracks and target measurements are present. The extensions of this algorithm such as Joint probability data associations (JPDA) have been designed to remedy constrains of the PDA.

The JPDA is practically identical with the PDA (9-12) except that the probabilities are computed for all of the tracks of all the targets and hence more hypotheses will be formed than in PDA. However, the JPDA computes the probabilities for all the tracks and the association algorithm have to be extended in that manner [12]. The extension of this algorithm has to govern the association probabilities of association of measurement to multiple tracks. It is

governed by Gaussian likelihood function of assignment of measurement i to the track j defined as:

$$g_{ij} = \frac{e^{\frac{-d_{ij}^2}{2}}}{(2\pi)^{m/2} \sqrt{|S_{ij}|}}.$$ (12)

Then the probabilities of all of the possible associations are computed, combined and used to update the track as in PDA (8-11).

The PDA and JPDA data association methods are the subject of intensive research and many modifications of these algorithms were developed. Some of the modifications are basically combination of PDA (JPDA) with the NN data association such as the NNPDA [18] and NNJPDA [19]. The other modifications are combination of PDA or JPDA with different tracking filters and the review can be found in [6], [20].

2.3 Multiple hypotheses tracking

The multiple hypotheses tracking (MHT) is generalization of the JPDA algorithm [13]. Like in JPDA an incoming measurement produces new hypotheses. For each new measurement three possible hypotheses can be formed. First hypothesis represents that the measurement is false alarm (e.g. is clutter originated), second case is that the measurement is originated from target and the third, the measurement originates from new target and hence new track for new target can be formed. The difference between the JPDA and MHT is that in the JPDA the association probabilities are computed only for the measurement (or set of measurements) received at one time instant e.g. in time instant *k*. In MHT the multiple different data association possibilities are maintained in the form of the hypothesis tree, which governs associations probability in multiple time instants [14]. Fig. 2 shows how the hypotheses are generated at every consecutive time instant. In this simplified case at time instant *k* only one track is present. At time *k+1* for new measurement two hypotheses are formed that the measurement should be associated with track 1 or that new track is formed.
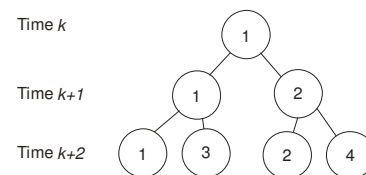
Fig. 2. MHT hypothesis tree

The evaluation of association probabilities is computed by using the track state estimation with covariance matrix and probability of association of new measurement with the existing track. The hypotheses (associations) which have low probability are cut off (pruned) from the tree. With this particular case of MHT, the hypothesis-oriented MHT, the computational complexity would grow exponentially with

the growth of number of hypotheses with new set of measurements. With the second case the target-oriented MHT the hypothesis tree is not maintained for all the time instants [15].

Since the MHT is the data association method that allows creating a new track based on the hypothesis tree it has been intensively studied. The review of MHT and its modifications can be found in [21].

3. Conclusion

In this paper a brief description of data association methods was presented. First the general principle and use of data association for radar sensor network was described. For the tracking of multiple targets with Kalman filter three basic methods of data association were outlined. It was found that use of particular one strongly depends on the computational capability of the radar system and on the dynamics of the tracking target. In future we would like to focus on the data association methods suited for different tracking filters.

Acknowledgements

References

[1] KMIOTEK, P.; RRUICHEK, Y., A fixed size assumption based data association method for coalescing objects tracking using a laser scanner sensor, Vehicular Electronics and Safety (ICVES), 2009 IEEE International Conference on , vol., no., pp.62-67, 11-12 Nov. 2009

[2] FOLSTER, F.; ROHLING, H., Data association and tracking for automotive radar networks, Intelligent Transportation Systems, IEEE Transactions on , vol.6, no.4, pp. 370- 377, Dec. 2005

[3] RASMUSSEN, C.; HAGER, G.D., Probabilistic data association methods for tracking complex visual objects, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.23, no.6, pp.560-576, Jun 2001

[4] NIEHSEN, W.; R. B. GmbH., Information fusion based on fast covariance intersection filtering. Fusion 2002, 2:901–904, August 2002.

[5] KOLAWOLE, M., Radar Systems, Peak Detection and Tracking. Newnes, April 2003.

[6] G. W. Ng, Intelligent Systems Fusion, Tracking and Control, RESEARCH STUDIES PRESS LTD. Baldock, Hertfordshire, England 2003

[7] HWANG, I., et.al., Multiple-target tracking and identity management, Sensors, 2003. Proceedings of IEEE , vol.1, no., pp. 36- 41 Vol.1, 22-24 Oct. 2003

[8] BLACKMAN S., POPULI. R., Design and Analysis of Modern Tracking Systems, Artech House, 1999.

[9] KONSTATNINOVA, P. UDVAREV, A., SEMERDIJEV, T., A Study of a Target Tracking Algorithm Using Global Nearest Neighbor Approach,

[10] ROECKER, J.A.; A class of near optimal JPDA algorithms, Aerospace and Electronic Systems, IEEE Transactions on , vol.30, no.2, pp.504-510, Apr 1994

[11] KYUNGMIN L; SHIN, V., Distributed fusion of local probability data association filters in multi-sensor environment, Computational Intelligence in Robotics and Automation (CIRA), 2009 IEEE International Symposium on , vol., no., pp.568-573, 15-18 Dec. 2009

[12] LIGGINS, M.E. II, CHEE-YEE Ch., KADAR, I., ALFORD, M.G.; VANNICOLA, V.; THOMOPOULOS, S., Distributed fusion architectures and algorithms for target tracking, Proceedings of the IEEE , vol.85, no.1, pp. 95- 107, Jan. 1997

[13] SANGHYUN CH, WOLF, M., BURDICK, J.W., An MHT algorithm for UWB radar-based multiple human target tracking, Ultra-Wideband, 2009. ICUWB 2009. IEEE International Conference on , vol., no., pp.459-463, 9-11 Sept. 2009

[14] CHENOUARD, N., BLOCH, I., OLIVIO-MARIN, J.-C., , Multiple hypothesis tracking in cluttered condition, Image Processing (ICIP), 2009 16th IEEE International Conference on , vol., no., pp.3621-3624, 7-10 Nov. 2009

[15] GIANNOPOULOS, E., STREIT, R., SWASZEK, P., Probabilistic multi-hypothesis tracking in a multi-sensor, multi-target environment, Data Fusion Symposium, 1996. ADFS '96., First Australian , vol., no., pp.184-189, 21-22 Nov 1996

[16] LI, X. R., BAR-SHALOM, Y., Theoretical analysis and performance prediction of tracking in clutter with strongest neighbor filters. In Proceedings of the 34th IEEE Conference on Decision and Control, volume 3, pages 2758–2763. IEEE, 13-15 December 1995.

[17] LI, X.R., XIAORONG Z., PSNF: a refined strongest neighbor filter for tracking in clutter ," Decision and Control, 1996., Proceedings of the 35th IEEE , vol.3, no., pp.2557-2562 vol.3, 11-13 Dec 1996

[18] FITZGERALD, R., Development of practical JPDA logic for multitarget tracking by microprocessor. In Multitarget-Multisensor tracking: Advanced Applications, pp. 1–23. Artech House, 1990.

[19] HELMICK, R. IMM estimator with nearest-neighbor joint probabilistic data association, in Multitarget-Multisensor Tracking: Applications and Advances, vol. III, Y. Bar-Shalom and D. Blair, Eds. Norwood, MA: Artech House, 2000, ch. 3, pp. 161–198.

[20] BLACKMAN, S.S., , Multiple hypothesis tracking for multiple target tracking, Aerospace and Electronic Systems Magazine, IEEE , vol.19, no.1, pp.5-18, Jan. 2004

[21] PULFORD, G.W., Taxonomy of multiple target tracking methods, Radar, Sonar and Navigation, IEE Proceedings - , vol.152, no.5, pp. 291- 304, October 2005

Dielectric properties of natural ester

[1]*Lýdia DEDINSKÁ*, [2]*Milan KVAKOVSKÝ*, [3]*Vieroslava SKLENÁROVÁ*

[1]Dept. of Electrical power enginnering, FEI TU of Košice, Slovak Republic
[2]Dept. of Electrical power enginnering, FEI TU of Košice, Slovak Republic
[3]Dept. of Electrical power enginnering, FEI TU of Košice, Slovak Republic

[1]lydia.dedinska@tuke.sk, [2]milan.kvakovsky@tuke.sk, [3]vieroslava.cackova@tuke.sk

Abstract—**Electroinsulating materials are important and expensive parts of every electrotechnical equipment. One of the most common used electroinsulating materials are electroinsulating oils. Dielectric properties of insulating oils will be described in this paper. The aim of this work was to investigate the electrical properties of vegetable oils and to compare them with the most commonly used insulating liquids.**

Keywords—**liquid dielectric, natural ester, mineral oil, electrical properties.**

I. INTRODUCTION

For over one-hundred years, petroleum-based mineral oils purified to "transformer oil grade" have been used in liquid-filled transformers. Synthetic hydrocarbon fluids, silicone, and ester fluids were introduced in the latter half of the twentieth century, but their use is limited to distribution transformers. Several billion liters of transformer oil are used in transformers worldwide. [1]

The popularity of mineral transformer oil is due to availability and low cost, as well as being an excellent dielectric and cooling medium. Ever since the world oil reserves were tapped in the 1940s, petroleum products have become widely available. Petroleum-based products are so vital in today's world that we cannot imagine a time we may not have them easily available. Transformers and other oil-filled electrical equipment use only a tiny fraction of the total petroleum consumption, yet even this fraction is almost irreplaceable. [1]

II. ELECTRICAL PROPERTIES OF LIQUID DIELECTRIC

A. Insulation resistance

Dependence of insulation resistance on time can be measured after connecting the DC voltage U [2]:

$$R(t) = \frac{U}{I(t)}.$$

(1)

where
R - is the insultanion resistance [Ω],
U - is the applied DC voltage [V],
I - is the current following thought of the measured sample [A]

In practice it is measured sufficient only two points R_{15} and R_{60} - resistance at 15 s. and 60 s., after connecting the voltage. The one minute polarization index p_{i1} implies of the time dependence over the insulation resistance and is characterized the state of insulation (the presence of moisture in the insulation). It is assessed according to the relationship [2]:

$$p_{i1} = \frac{R_{60}}{R_{15}}.$$

(2)

where
R_{15} - is the resistance at 15 s. [Ω]
R_{60} - is the resistance at 60 s. [Ω]

The real isolation has unobtainable a value 1 even good isolation, because in reality is the structure of insulation non-homogeneous, and this inhomogeneity be maintained in ideal desertification. [2]

According to the values of insulation resistance can be assessed the state degradation of insulation. It leads to a significant reduction insulation resistance. [2]

B. Dielectric loss factor

The dielectric loss factor includes to the most important parameters which is related to dielectric loss and points to quality electric - insulating material. Coefficient of dielectric loss factor tgδ is defined as tangent of the angle between phase displacement of real dielectric and ideal dielectric. The value of loss factor depends on temperature, frequency and intensity of the acting electric field. The increasing value of dielectric loss factor tg δ points to reduce the quality of insulation. [2]

Dependence of loss factor on the voltage

The value of loss factor depends on voltage. Increasing of voltage connected to insulation will cause changes of some processes, which have influences on the value of dielectric losses. [2] We are able to observe these phenomena mostly in case when above specific voltage level, the different mechanism of losses will start which is caused by ionizing processes. By measuring of dependency tgδ = f(U) in periodical intervals we are able to consider changes which occurred during the machine operation.
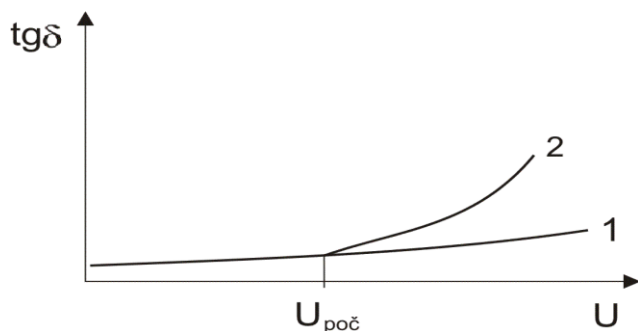
Fig. 1. *Dependence of loss factor on the voltage. [2]*

In the figure (Fig. 1) curve 1 is for the quality insulation system, where the value of loss factor depends on voltage should be constant. Curve 2 is for the insulation system with ionizing processes, and the value of loss factor is increased.

Dependence of loss factor on the temperature

The loss factor tgδ of insulation is occurred on temperature. Losses related to the electrical conductivity of insulation are increased monotonically with increasing temperature (curve 1, fig.2) Polarization losses (of the dipole) are also reported the maxim for homogeneous insulator in depending at the temperature (curve 1 fig.2). In the figure (Fig. 2) curve 3, is displayed temperature dependence of tgδ of insulation that dielectric loss are affected by conductivity and polarization. [2]
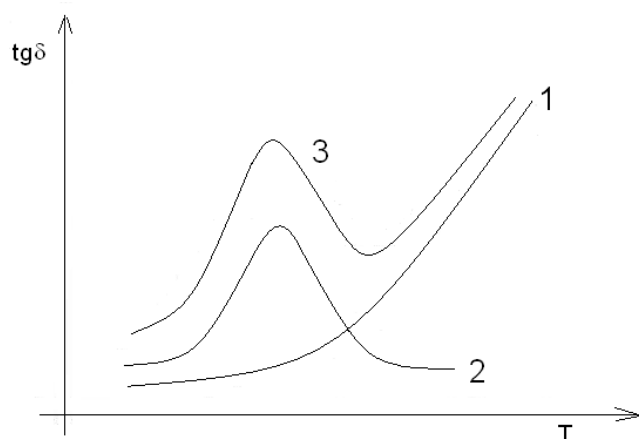
Fig. 2. *Dependence of loss factor on the temperature. [2]*

Permittivity

Permittivity ε_r is defined as a ratio of capacity of electrodes in a space fully filled by appurtenant dielectric and capacity of the same configured electrodes in the vacuum (practically in the air too, because the relative permittivity of the air is 1,00058). [2]

Permittivity specifies how many times energy functioning on a charge in the appurtenant space is lower than in the vacuum:

$$\varepsilon_r = \frac{C_x}{C_0} \qquad (3)$$

where: ε_r - is the dielectric permittivity [-],
          C_x - is the capacity of the full oil tank, the
              capacity of the measured sample [pF],

C_0 - is the capacity of the empty oil tank, the
          geometric capacity of the sample [pF]
Permittivity depends on the polarization processes, temperature, and frequency.

C. Breakdown voltage and breakdown strength

Breakdown voltage presents degree of ability of oil to resist an electric stress. It is the minimal voltage value, which causes rise of electric conductivity to the level that causes an electric breakdown. High value of current abounding through breakdown area , causes mechanical, thermal and chemical processes that change dielectric characteristics. These changes are so significant, that dielectric is not able to completely return to its regular condition, because of solid particles and chemical compounds in liquid or vapor consistency, which come to being by electric discharge activity. Free water, gas bubbles and solid particles are inclined to migrate into the areas with enhanced stress and it causes lower value of breakdown voltage. For this fact, the breakdown voltage can be used as an indicator of oil pollution. [3]

Electric breakdown strength is one of the basic qualitative characteristics of dielectrics in addition to polarization and dielectric loss. In electric field, dielectric keeps its insulating characteristic only up to specific value of electric field intensity. After reaching this boundary (critical) field intensity, resistance of dielectric decreases rapidly to resistance level of conductive materials. [7]

In case of a homogeneous electric field, field intensity is the same at full length of breakdown trajectory, therefore electric breakdown strength can be calculated using the pattern:

$$E_b = \frac{U_b}{d} \qquad (4)$$

where:
          E_b - is the dielectric breakdown strength [kV/mm],
          U_b - is the breakdown voltage [kV],
          d   - is the interelectrode distance [mm]

Electric breakdown strength is parameter most common used for oil quality arbitration. Ranking electric breakdown strength correctly in liquid dielectric is more difficulty as in solid or vapour dielectrics. The main reason is fact that breakdown process in insulant liquid and value of breakdown voltage depends on number of random factors. Among the most important belongs water content, solid impurity content, electrode configuration and applied voltage period.

III. MEASURING RESULTS

Investigations were focused on the measurement of the dissipation factor of natural ester and comparison to mineral based transformer oil and silicon oil. As vegetable oil was used cloza oil. As transformer oil was used inhibit oil ITO100. As silicon oil was used LUKOSIOL M 200.

From fig.3 for mineral oil ITO 100 is visible, that value of the one minute polarization index p_{i1} drops down to temperatures.
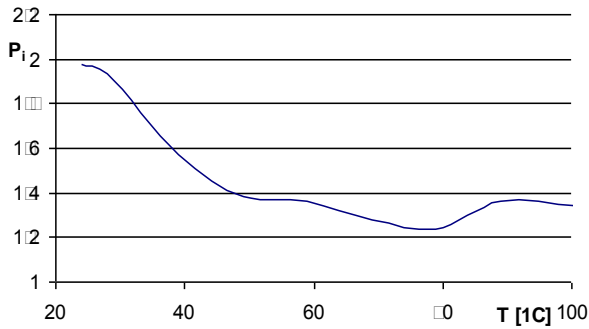
Fig. 3. *Temperature dependence of the polarization index of mineral oil ITO 100*

Fig. 4. Voltage dependence of loss factor tg δ for T = 20°C for cloza oil, ITO 100 and Lukosiol M 200 samples

Fig. 5. Voltage dependence of loss factor tg δ for T = 50°C for colza oil, ITO 100 and Lukosiol M 200 samples

The voltage dependence is the least significant for natural ester colza oil, which retained constant although high values of

dielectric loss, for all the voltage levels (Fig. 4, 5). The dependence of dielectric loss on voltage levels for mineral and silicone oil is minor for the higher values of voltage and temperature (Fig. 6).

Fig. 6. Temperature dependence of loss factor tg δ for U = 2kV for cloza oil, ITO 100 and Lukosiol M 200 samples

Fig. 7. Temperature dependence of permitivitty ε for U = 2kV for cloza oil, ITO 100 and Lukosiol M 200 samples

From the results (Fig. 7) is able to see, that the permittivity values declines with increasing temperature in all of the oil samples, whereby the trend of the regress is approximately constant. Very high and the highest value of permittivity reached natural ester colza oil (ε_r = 3,2), what is really desirable for combined oil – paper insulation, which is often used in power transformers. Paper insulation reaches the permittivity value ε_r = 4. From the view of temperature dependence, the biggest disadvantage of natural ester colza oil is expressive dependence and high values of loss factor by the temperature values advanced than operative. From this point of view, the best results reached silicone oil. The dependence of the dielectric loss on temperature is just of small account.

As shown on the graph (Fig. 7), it is no possible to say strictly, which oil sample is the best, or the worst from perspective of electric breakdown strength. The results are not clear, mainly for the smallest interelectrode distances. For interelectrode distance 0.3mm, natural ester colza oil and silicone oil Lukosiol M200 reached the highest values of breakdown strength. Mineral oil ITO 100 proves its qualities for bigger interelectrode distances. Natural ester colza oil did not overreach values of breakdown strength of mineral oil, but

reached strikingly the very high and comparable values of the results.
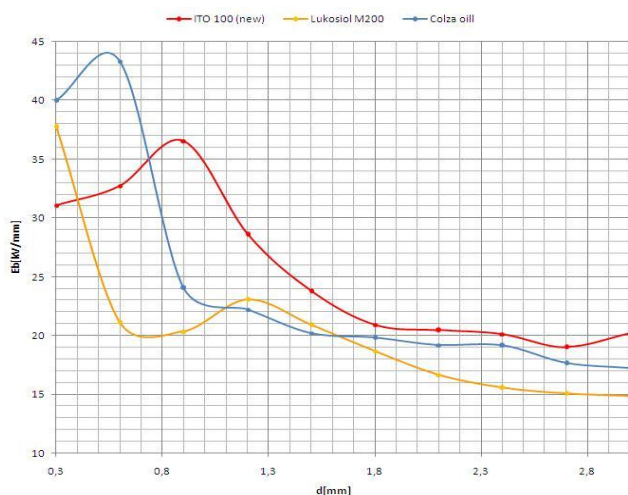
Fig. 8. Comparison of $E_p = f(d)$ characteristic for cloza oil, ITO 100 and Lukosiol M200 samples

IV. CONCLUSION

The aim of this work was to compare the dependence of dissipation factor, permittivity and electrical breakdown strength for liquid dielectrics. Natural oils: transformer oil ITO 100 and silicon oil Lukosiol M 200 were compared with natural ester fluid: colza oil.

It is visible from charts, that the value of dissipation factor tg δ of liquid insulators rises with temperature. It was also shown, that tg δ rises only very slowly with the value of applying voltage. The dissipation factor decreases by lower values of voltage up to 0.6kV for transformer oil ITO 100 and silicon oil Lukosiol M200. It remains constant for higher values of voltage. The greatest advantage of natural ester colza oil is the high value of permittivity, which is similar to the permittivity value of paper. Mineral oil ITO 100 had the lowest values of permittivity. The values of the electric breakdown strength of the natural ester colza oil are comparable with the values of the mineral oil.

ACKNOWLEDGMENT

REFERENCES

[1] Oommen, T.V.: Vegetable Oils for Liquid-filled Transformers. In.: Magazine Electrical Insulation IEEE, volume 18, number 1, january/February 2002, pp.6-11
[2] Kolcunová, Iriada: Diagnostika elektrických strojov, Technická univerzita Košice, 2006, ISBN 80-8073-550-.
[3] Poljak, František: Dielektriká izolanty, ALFA, 1983
[4] Tichý, V. - Pallo, V. - Mašek, V.: Transformátorový olej. , STNL, Bratislava, 1962 .
[5] DOLNÍK, B. – KURIMSKÝ, J.:The Change of Relaxation Current Pattern Due to Forced Thermal Aging of Sintered Ceramics. Electric Power Engineering 2010: proceedings of the 11th international scientific conference, 4.-6. 5. 2010, Hotel Santon, Brno, Czech Republic. - Brno, University of Technology, 2010. ISBN 978-80-214-4094-4. P. 739-742.
[6] Marton, K., Kurimský, J.: Diagnostika transformátorov Publikácia TUKE: Starnutie elektroizolačných systémov. č. 2 (2007), s. 3-13. ISSN1337-0103 Dostupné na internete: <http://jeen.fei.tuke.sk/jeen2/index.php/JSES/article/viewFile/88/90>
[7] Transformer Fire Protection, Facilities Instructions, Standards, and Techniques Volume 3-32, www.usbr.gov/power/data/fist/fist3_32/fist3-32.pdf
[8] PETRÁŠ, Jaroslav: Partial discharge acoustic emission signal analysis. In: Przeglad Elektrotechniczny Konferencje vol.5, no. 3 (2007), pp.177-179, ISSN 1731-6103
[9] TKÁČ, Ján - KURIMSKÝ, Juraj - DOLNÍK, Bystrík: Laboratórny výskum výbojovej činnosti v modeloch s kombinovanou izoláciou papier-olej. In:Starnutie elektroizolačných systémov. č. 4 (2008), s. 18-20. ISSN1337-0103 Internet: http://web.tuke.sk/fei-kee/jses/uploads/File/jses-04-2008-07.pdf

Dynamic Emulation of Mechanical Loads – an Overview of Speed Control Structures

[1]*Karol KYSLAN*, [2]*Matúš HRIC*, [3]*Peter NGUYEN*

[1,2,3] Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]karol.kyslan@tuke.sk, [2]matus.hric@tuke.sk, [3]peter.nguyen@tuke.sk

Abstract — **The paper presents an overview of control structures used in dynamical emulation of mechanical loads. Only speed control structures are presented. The structures allow applying control algorithms that enable testing a drive without real mechanical load.**

Keywords — **dynamic emulation, drive control, test bench**

I. INTRODUCTION

In the field of computer science term *emulation* was used by the IBM Corporation for the first time in order to term a special kind of software. This software was able to imitate certain platform behaviour by running on another platform. In general, emulation is the imitation of the system or the part of the system by another platform or technical device so that imitating system behaves the same way as imitated system would. For the same input data of imitated and imitating system the output data have to be the same. The term for the imitating system is *emulator* and the term for this imitating technique is *emulation*.

In recent years methods of emulation are being used in design and validation of control algorithms for variable position, speed and torque drives used in mechatronic applications. These algorithms can be tested with the use of computer simulations. Standard verification of control structures by the mathematical modelling is improved when using one or several actual devices instead of their simulation models. The other parts of the process are simulated in appropriate real-time system. Hardware running real-time system and equipped by DAQ interface boards in conjunction with actual devices creates well-known HIL simulation, which is nowadays more and more used to develop new structures and components in many fields.

It is of great benefit to use the emulation technique for the device which is not available for the algorithm testing. For example high power drive, multimotor drive, etc. are rarely available in laboratory facilities. Furthermore, in economic and space terms, it is not practical to have different kinds of mechanical loads in the laboratory. More convenient approach is to emulate mechanical loads electrically. The most profitable solution is to use two electrical drives with machines connected in common shaft as presented in Fig. 1. For the one of the drives control algorithms will be developed. It is termed *drive under test*. Drive under test is loaded by using torque controlled *load machine*, also known as *dynamometer*. The torque of the dynamometer is controlled

such that it creates the same shaft torque on the drive under test as the desired mechanical load would. The technique is widely known as *load emulation*.

The dynamometers have been mostly used to perform static load tests of electrical machines [1]. Another approach is the use intended for emulation of torque-speed characteristis enabling the testing of machines under steady state or slowly changing conditions. This method is known as the static emulation [2], [3]. The upgrade of the method is the emulation of the simulated load under the open-loop conditions [4]. Methods mentioned above are not sufficient for testing of variable speed and torque drives in cases with dominant nonlinear and dynamic effects like high changing inertia, rapid changes of speed, etc. Thus, methods with closed-loop conditions are being developed.

The very first method of dynamic emulation in closed-loop uses inverse dynamic model of the emulated load, which brings many problems, as sampling effects [5], calculating the derivative terms in the inverse transfer functions [6] and violation of open loop pole-zero structure. It was shown in [6] that if emulated inertia exceeds twice that of real inertia, system is unstable. Because of these limitations this approach is not appropriate to be used in control of load machine and new techniques turned up to be required. In this paper three different approaches of dynamic emulation are briefly described.
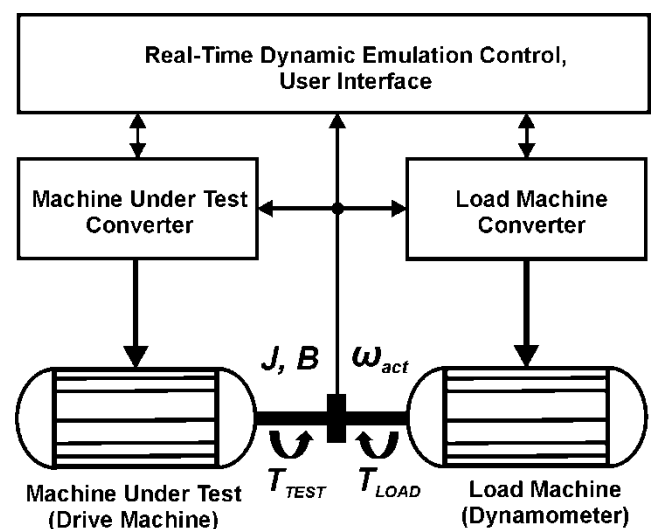
Fig. 1. The arrangement of the experimental system for dynamic emulation of mechanical loads
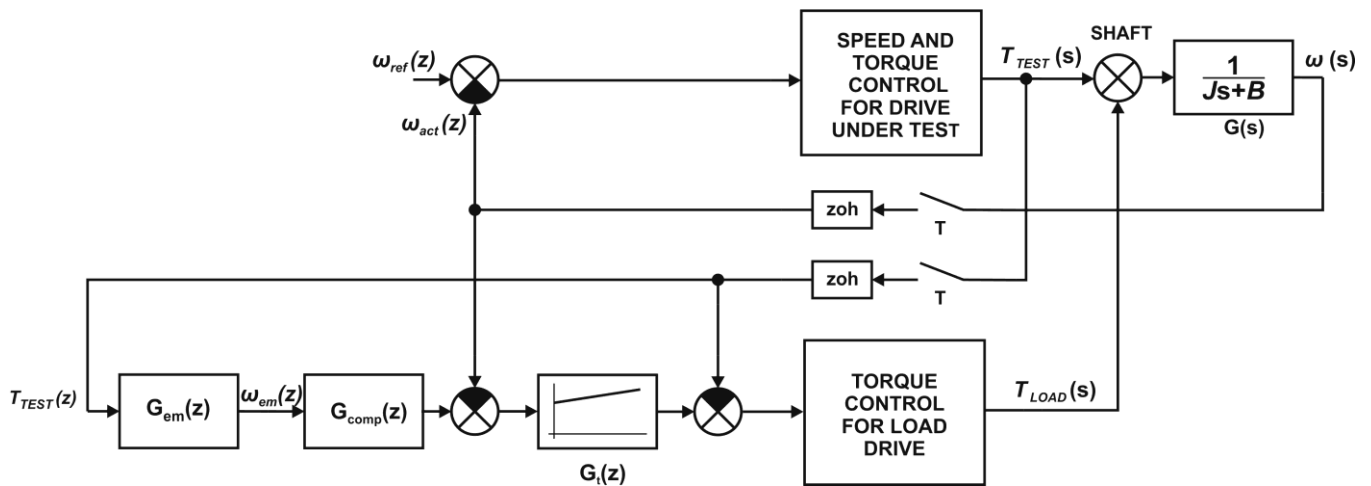
Fig. 2. Overall control structure of dynamic emulation using compensator and feedback controller

II. DYNAMIC EMULATION USING COMPENSATOR

The control structure is given in Fig. 2. The torque of the load machine T_{LOAD} is added to the torque of the drive under test T_{TEST}. J is the inertia of both machines and B is the viscous friction coefficient of the test bench. G_{em} (z) is the load to be emulated like unwinder, electric vehicle, flexible shaft, etc. Its parameters are given by the user and it is used to derive an ideal angular velocity demand ω_{em} (z) at which we wish to go if T_{TEST} (z) is applied to the load being emulated. Block G_{comp} (z) cancels all parts of open loop except the emulated load transfer function in order to achieve:

$$G_{em}(z) = \frac{\omega(z)}{T_{TEST}(z)}, \qquad (1)$$

what is the essential equation of dynamic emulation [8]. Block $G_t(z)$ is the speed-tracking loop controller and generates reference torque for the load drive. The delays introduced by torque control loops are ignored in derivation of final closed-loop transfer function. The benefits of this approach are that inverse model of emulated load is not needed and the dynamics of emulated load is preserved. The drawbacks of this approach are that transfer function of mechanical bench $G(s)$ is supposed as linear and additional delays caused by dynamics of speed-tracking controller $G_t(z)$ included in transfer function of $G_{comp}(z)$ are occurred [13]. More information about method can be found in [6]-[9] and position control design can be found in [10].
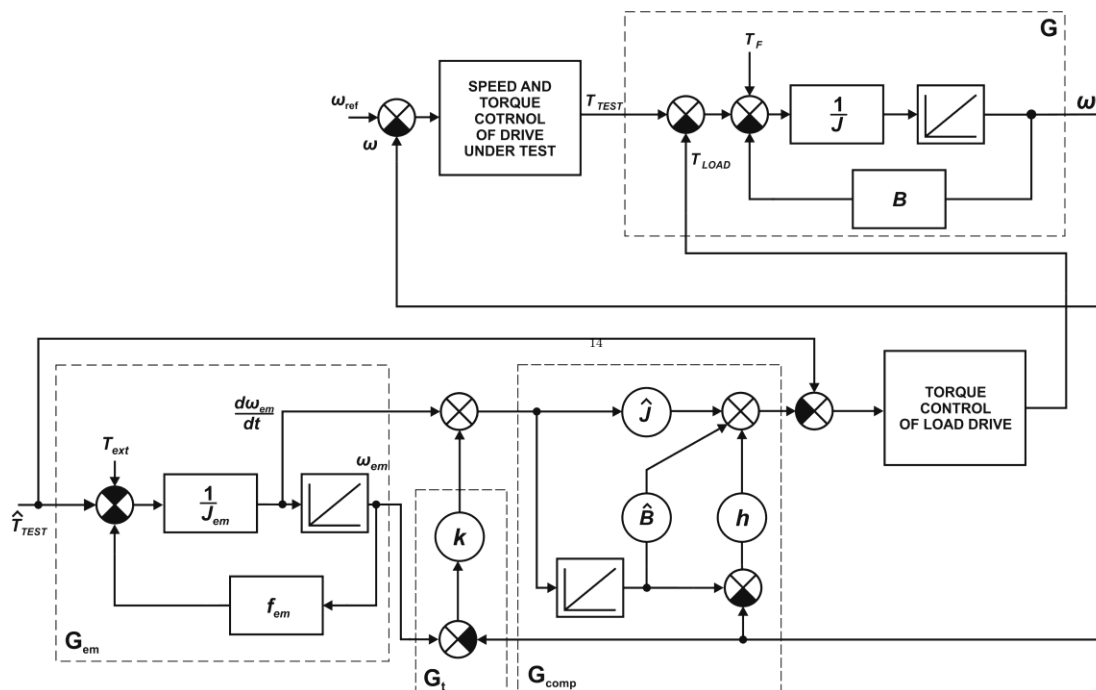
Fig. 3. Overall structure of dynamic emulation based on nonlinear control

III. DYNAMIC EMULATION BASED ON NONLINEAR CONTROL

This approach was developed for mechanical benches with some nonlinear features. The delays caused by dynamics of speed-tracking controller are not included as in the previous approach because of different compensation method used. The control structure is given by Fig. 3. Ideal angular velocity demand is again derived from the load to be emulated G_{em}. But in addition a signal of angular acceleration:

$$\varepsilon_{em} = \frac{d\omega_{em}}{dt}, \tag{2}$$

must be available in order to use it in the compensator G_{comp}. The unmodelled nonlinear dynamics of mechanical bench is presented as torque input T_f. J_{em} is the emulated model inertia, B_{em} is the emulated model viscous friction coefficient and f_{em} presents the emulated model friction as well as any other contribution of the load like gravity, centrifugal forces, etc. T_{ext} presents an additional load torque input applied to emulated model. $\hat{T}_{TEST}$ is the estimated value of the torque obtained either from measurement or with the use of an observer. In vector controlled ac machines well-known current model observer can be implemented. $\hat{J}$ and $\hat{B}$ presents the estimated values of mechanical bench's inertia and viscous friction, h presents the gain of compensator and k is the gain of feedback state controller. Nonlinear control of the system consists of cancelling the linear dynamics of mechanical bench by estimated values of $\hat{J}$ and $\hat{B}$ and cancelling the nonlinear dynamics of bench by PI estimator. After that the system is to be assumed a linear and controlled by linear state controller.

The benefit of this approach is that mechanical rig with some nonlinear effects can be used. However, if the features of mechanical rig are not well known, the operation is limited to certain bandwidth. More information about the approach can be found in [11], [12] and the position control design can be found in [13].

IV. DYNAMIC EMULATION BASED ON PRE-CONTROL

The control structure of this approach is given by the Fig. 3. In this structure all variables are in p.u. formulation and marked in round parentheses. The transfer functions G_{TEST}, G_{LOAD}, and G_{MOD} present equivalent transfer functions of the torque-, test-, load-, and emulated drive circuits. The idea of the emulator design consists in the following: the p.u. value of the torque reference (T_{set}) is fed both into the tested drive and into the model. To get the same dynamic responses, both drives must satisfy the following conditions:

$$\omega_{TEST} = \omega_{MOD} \tag{3}$$

$$\dot{\omega}_{TEST} = \dot{\omega}_{MOD} \tag{4}$$

The reference torque generated by the emulation control satisfying the conditions (3) and (4) is given by (5).

$$(T_{REF\ LOAD}) = \frac{G_{TEST}}{G_{LOAD}}(T_{REF}) - \frac{J_{TEST}}{J_{MOD}}\frac{G_{MOD}}{G_{LOAD}}(T_{REF}) + \frac{J_{TEST}}{J_{MOD}}\frac{1}{G_{LOAD}}(T_{EXT}) \tag{5}$$

Equation (5) presents the drive under test pre-control. This enables to use an arbitrary shape of the load torque and moment of inertia providing its dynamics is lower than load drive torque loop dynamics. Although the emulator forces the dynamics to the tested drive, it is done through the load torque and not through the control circuits like it is used in forced dynamics systems [17]. With p.u. variables used it is possible to test high-power systems in laboratory conditions with the test bench of low-power. More information about the approach with simulation and experimental results can be found in [14]-[16].
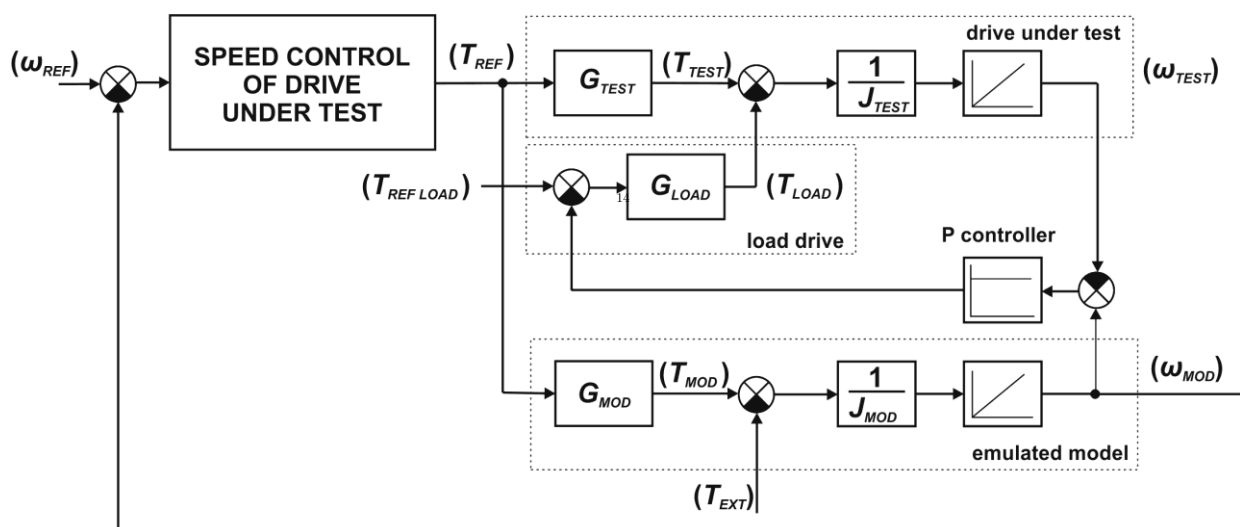
Fig. 4. Overall structure of dynamic emulation based on pre-control

V. Conclusion

In this paper essential control structures for dynamic emulation of mechanical load are briefly described and their characteristics are presented. Without the need for the real mechanical load, presented methods enables the development of control algorithms for various mechatronic applications driven by electric machines.

Dynamic emulation using compensator presents excellent results, but it was developed the test rig, which is considered to have a dynamics very close to linear. This drawback is eliminated by using of dynamic emulation based of nonlinear control. Dynamic emulation based on pre-control the only one enables testing the control algorithm for high-power drives in laboratory environment using low-power drive.

Acknowledgment

References

[1] C.R. Wasko, "A universal AC dynamometer for testing motor drive systems," in *Conf. Rec. IEEE-IAS Annu. Meeting*, pp 409-412, 1987.

[2] Collins, E.R., Jr.; Huang, Y.: „A programmable dynamometer for testing rotating machinery using a three-phase induction machine," *Energy Conversion, IEEE Transactions on* , vol.9, no.3, pp.521-527, Sep 1994

[3] Newton, R.W., Betz, R.E., Penfold, H.B.: „Emulating dynamic load characteristics using a dynamic dynamometer , " *Power Electronics and Drive Systems, 1995., Proceedings of 1995 International Conference on* , vol., no., pp.465-470 vol.1, 21-24 Feb 1995

[4] Sandholdt, P., Ritchie, E., Pedersen, J.K., Betz, R.E.: „A dynamometer performing dynamical emulation of loads with nonlinear friction, "Industrial *Electronics, 1996. ISIE '96., Proceedings of the IEEE International Symposium on* , vol.2, no., pp.873-878 vol.2, 17-20 Jun 1996

[5] Akpolat, Z.H., Asher, G.M., Clare, J.C.: „Dynamic emulation of mechanical loads using a vector-controlled induction motor-generator set, " *Industrial Electronics, IEEE Transactions on*, vol.46, no.2, pp.370-379, Apr 1999

[6] Akpolat, Z.H., Asher, G.M., Clare, J.C.: „Experimental dynamometer emulation of nonlinear mechanical loads, " *Industry Applications, IEEE Transactions on*, vol.35, no.6, pp.1367-1373, Nov/Dec 1999

[7] Arellano-Padilla, J., Asher, G.M., Sumner, M.: „Control of an AC Dynamometer for Dynamic Emulation of Mechanical Loads With Stiff and Flexible Shafts, "*Industrial Electronics, IEEE Transactions on* , vol.53, no.4, pp.1250-1260, June 2006

[8] Žalman, M., Macko, R.; „Design and realization of programmable emulator of mechanical loads," *16th IFAC World Congress*, Volume 16, Czech Republic, July 2005

[9] Macko R., Žalman M., Uhríček M.: „Programmable Emulator of Mechanical Loads from Motors", *AT&P journal No 2*, 2005, pp. 92-95.

[10] Akpolat, Z.H., Asher G.M., Padilla, J.E.: „A test bed for the experimental validation of position control algorithms", *Control Engineering Practice, Special Section on Emerging Technologies for Active Noise and Vibration Control Systems*, Volume 12, August 2004, Pages 933-943, ISSN 0967-0661

[11] Rodic, M., Jezernik, K., Trlep, M.: „Dynamic emulation of mechanical loads - approach based on nonlinear control," *Power Electronics and Applications, 2005 European Conference on* , vol., no., pp.10 pp.-P.10, 0-0 0

[12] Rodic, M., Jezernik, K., Trlep, M.: „Dynamic emulation of mechanical loads: an advanced approach," *Electric Power Applications, IEE Proceedings* - , vol.153, no.2, pp. 159- 166, 2 March 2006

[13] Rodic, M., Jezernik, K., Trlep, M.: „Dynamic emulation of mechanical loads — position control approach, " *Power Electronics and Motion Control Conference (EPE/PEMC), 2010 14th International* , vol., no., pp.S10-23-S10-30, 6-8 Sept. 2010

[14] Ďurovský, F., Fetyko J., Fedák V.: „Testing of Drives by a Load Torque Emulator". *Automatizácia a riadenie v teórii a praxi* No. 5, 2009, pp. 150/1–150/6, www.strojarstvo.sk.

[15] Ďurovský, F. et al.: „Emulation of a winder drive". In: *EDPE 2009: Abstracts & CD Proceedings of 15th International Conference on Electrical Drives and Power Electronics*, 4th Joint Croatia-Slovakia Conference: October 12-14, 2009, Dubrovnik, Croatia. Zagreb: KoREMA, 2009. 5 p. ISBN 978-953-6037-56-8

[16] Ďurovský, F. et al.: „Load Torque Emulator in Testing Power Converters for Small Vehicles". In: *XXIV. Int. Scientific Conference "microCAD,* 18.-20. 3. 2010, Section K: Electrotechnics and Electronics. University of Miskolc, 2010. P.19-26.

[17] Dodds, S.J., Vittek, J.: „Riadenie elektrických pohonov s vnútenou dynamikou", EDIS, vydavateľstvo Žilinskej univerzity, Žilina, 2003. ISBN 80-8070-087-7

[18] Kyslan, K.: „Load torque emulator based on industrial converters". In: *SCYR 2010 : 10th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice*: proceedings from conference: May 19th, 2010, Košice, Slovakia. - Košice : FEI TU, 2010 S. 95-97. - ISBN 978-80-553-0423-6

Effect of Aging and Moisture on the Insulation System

[1]*Vieroslava SKLENÁROVÁ,* [2]*Lýdia DEDINSKÁ,* [3]*Milan KVAKOVSKÝ*

[1.] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
[2.] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
[3] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]vieroslava.cackova@tuke.sk, [2]lydia.dedinska@tuke.sk, [3]milan.kvakovsky@tuke.sk

Abstract— **This article discusses the aging of the insulation system and the influence of moisture content on the aging transformer oil during IRC Analysis.**

Keywords—**Aging, moisture, temperature, insulation system, transformer oil.**

I. INTRODUCTION

The effectiveness use of electrical equipment is necessary to know the rules of aging of insulation systems. Life of electrical equipment is mostly determined by the lifetime of the insulation system. Power transformers have to the required electrical insulation properties and mechanical properties that match the voltage range and the expected mechanical stresses during short circuit events. Developments trends are determined by monitoring over changes of properties that are directly related to the electrophysikal structure of material. Anomalies in these process point to change the mechanism of aging, which is in most cases associated with the occurrence and development of fault insulation system. This disorder may have a different character.

Moisture content during aging in liquid and solid insulation reduces dielectric strength, accelerates the decay of cellulose and causes emission of bubbles at high temperatures. Equilibrium moisture content is based on thermodynamic equilibrium, where the migration of moisture molecules inside materials and the oil and cellulose is zero. Temperature change will always affect the migration of moisture inside the material and the cellulose and oil, because the solubility in water and oil sorption capacity of cellulosic materials depends on temperature.

The ability of oil-soluble water increases particularly aging

II. AGING OF INSULATION

A. Insulation

Insulation of electrical equipment is exposed to electrical stress during the whole services. Experimental measurements and laboratory tests are clearly showed the great influence of the applied voltage on the aging of insulation systems. Strength of electric field is increased over time and the aging of insulation is shorted. The classification of the impact of electrical stress on the electrical equipment is necessary to distinguish the impact of direct and alternating electric field. Insulating materials can be divided into two groups, according to impact of electric field to the material [1]:

- Thermoreaktive
- Thermoplastic

Thermoreaktive materials are characterized by the existence of a threshold below which material life is theoretically infinite, unless there is present another kind of stress. It is possible to define the relationship [4]:

$$\tau = H \frac{exp(-h.E)}{E - E_T} \qquad (1)$$

where: E - the electric stress (effective value of the applied ac voltage, respectively. voltage gradient of electric field in which the material is stressed)

E_T - the initial stress,

H, h - constants dependent on temperature
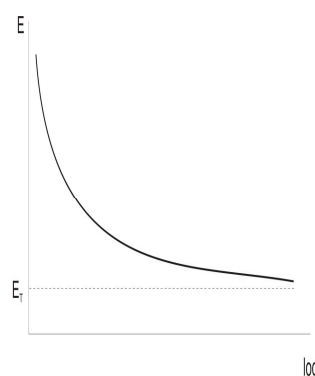
Fig. 1. Aging of thermoreaktive materials [1]

Thermoplastic materials are a group of materials that have been no observed threshold limiting aging. Of course, even in these materials were expected threshold electrical stress, but is so low that it is only in extreme long times beyond the technical use of the material [1].

The aging curve of thermoplastic materials is not precisely defined for thresholds and its stabilization was not observed even at very long times. The thermoplastic materials are described best by reciprocal square rule, which is also confirmed statistically [1].

$$\tau = C.E^{-n} \qquad (2)$$

where: C - constant dependent on temperature,

n – coefficient of voltage resistance dependent on temperature.

The coefficient n is important data about the quality material in terms of its resistance to voltage stress. The coefficient n represents slope of aging line. The coefficient n is more, the material is more durable.
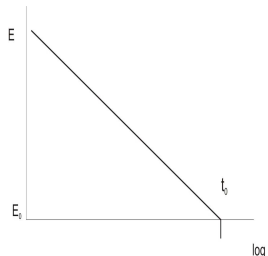
Fig. 2 Aging of thermoplastic materials [1]

B. Thermal stress

Thermal stress - aging is a gradual physical and chemical changes as sequence of chemical degradation reactions. The temperature is factor with essential importance for the operational capability of electrical equipment. All processes are intended to thermodynamic equilibrium. Operating property is affected by the effect of

- high or low ambient temperatures,
- increased in dielectric loss emergent in the isolation of electrical equipment,
- elevated temperatures above ambient temperature

As a result of heat stress is permanently overloading of isolation and reduction of its electrical insulating properties and its accelerated aging. Electrical insulation properties of gas and liquid insulators can be renewed (oil refining, oil purification etc.), but in solid insulators are changes irreversible.

The time that elapses before the limiting value when the insulation is no longer able to safely perform the prescribed function can be written according to Arrhenius [1] [2]:

$$\tau = A.exp\frac{B}{T} \qquad (3)$$

where: τ - time of aging

T - absolute temperature,

A, B - constants determined of activation energy of reaction.

Arrhenius relationship is based on the description of the dependency rate of chemical reactions from first-order temperature [3].

Analysis of processes are arising in isolation, while respecting temporal changes of the concentration of material components and their relationship to changes in other physical quantities, we can take from several angles, which will the basis form of Arrhenius theory.

File of processes in transformers oil leads to changes of properties (physical, chemical or electrical) in operating conditions. Changes can to reduce the required insulating properties of transformer oil for operational safety devices.

The aging process of transformer oil is affected of different impacts:

- the reducing of the mechanical strength in cellulose insulators as a result of acids attack incurred aging process
- the reducing the electrical strength of oil and insulating systems oil – paper as products due to aging
- deterioration of cooling as a result of settling of sludge in the oil canals, the increasing of temperature at 8 to 10 K is caused a doubling rate of aging (Montsingers rule)

Isothermic relaxation current analysis (IRC analysis) is enabled to analyze non-destructive aging of system. IRC analysis is allowed to define the elements of equivalent circuit and calculate the dielectric constant of aging A. It is necessary to take into account the time zone and the relevant spectrum, while the equivalent circuit elements must be independent of voltage and frequency.

Coefficient of aging A has been empirically determined to assess and quantify the replenishment.

if $\tau_3 \succ 3.\tau_2$

and $\tau_2 \succ 3.\tau_1$

then to coefficient of aging A apply

$$A = \frac{Q(\tau_3)}{Q(\tau_2)} = \frac{1 + \frac{a_2.\tau_2}{a_1.\tau_1}.(1 - e^{-\frac{t_3}{t_2}}) + \frac{a_3.\tau_3}{a_1.\tau_1}.(1 - \frac{1}{e})}{1 + \frac{a_2.\tau_2}{a_1.\tau_1}.(1 - \frac{1}{e}) + \frac{a_3.\tau_3}{a_1.\tau_1}.(1 - e^{-\frac{t_2}{t_3}})} \qquad (4)$$

where: a_i, τ_i - amplitudes of elementary streams and equivalent times in the i-th branch of the replacement model of the dielectric in the development of three RC members (obtained by calculation in the third approximation).

We made repeated measurements of new mineral oil MN by method of IRC analysis. The temperature is increasing gradually by step 10° C and monitoring the polarization processes the during the 1000 s. The first day we made measurements at 20 and 30° C, the second day measurements took place at 40 to 60° C, the third day for 70 and 80° C. We made the measurements at 90 and 100° C last day. Subsequently, we evaluated value of IRC analysis and approximated to third polarization processes. Next we calculated the aging factor A (Table 1) during relationship (4).

TABLE I
CALCULATE THE AGING FACTOR A FOR MN

	MN								
	24℃	30℃	40℃	50℃	60℃	72℃	81℃	91℃	101℃
I_{m311}	20788	27903	27553	28012	27400	20001	29706	23606	15975
I_{m321}	2375	3506	3125	3163	3078	2176	3278	2403	1587
I_{m331}	476	648	653	638	688	523	699	508	400
I_{m301}	51	69	68	70	66	72	70	68	50
τ_{311}	0,31	0,32	0,29	0,33	0,32	0,41	0,32	0,43	0,43
τ_{321}	3,04	3,17	2,67	3,72	3,32	4,1	3,48	5,47	4,41
τ_{331}	42	44,45	33,66	44,72	39,89	40,21	41,31	55,54	49,21
Q_{t3}	4,08	4,28	3,78	4,22	4,14	3,71	4,12	4,05	3,83
Q_{t2}	1,92	2,01	1,87	2,05	1,99	1,94	2,00	2,08	1,89
A	2,12	2,13	2,02	2,06	2,09	1,92	2,06	1,95	2,03

In the fig. (3) is evident that interruption of measurement is clearly affected the rate of aging A.
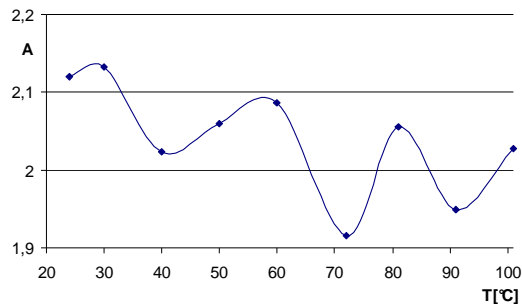
Fig. 3 Coefficient of aging A of MN

In our opinion, the measurement is occurred of access to atmospheric moisture to the sample discontinuation of. If the measurement sample of oil is gradually heated, then moisture getting rid continuation in part again. Our measurement results will affected, as is shown in Figure (1).

III. CONCLUSION

Water is a natural part of oil. In our laboratory conditions is difficult to ensure the elimination of access to atmospheric moisture. Our measurements are significantly demonstrated affects of moisture the measured results, and thereby evaluating the coefficient of aging A. We propose to implement a continuously increasing temperature from 20 to 100 ° C to exclude the impact of moisture measurements,. The second option is long-term heating oil (about 6 hours) for the temperature from 20 to 100° C. These methods and results will be evaluated in our other measurements.

REFERENCES

[1] MARTON, K.: "Diagnostika VN a VVN zariadení", ČSVTS, Stará Lesná, 1987
[2] SIMONI, L., MAZZANTI, G., MONTANARI, G. C., LEFEBRE, J.: "A general Multi-stress Life Model for Insulation Materials with or without Evidence for Threshholds", IEEE Transaction on Electrical Insulation, Volume 28, Number 3, 1993
[3] LELÁK, J.: "Starnutie, stanovenie zvyškovej životnosti a kvalifikácia elektrotechnických prvkov a zariadení", STU, Bratislava, 1996

Elastomagnetic sensor of pressure force output signal verification

Anna HODULÍKOVÁ

Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

anna.hodulikova@tuke.sk

Abstract— **The presented article deals with verification of output signal (voltage) of elastomagnetic sensor (EMS). The goal of the verification was to compare the results gained from magnetic field simulation in the ferromagnetic sensor´s core (with numerical calculations) with the experimental results gained by measuring. The simulations were realised in COSMOS/EMS software environment.**

Keywords— **simulation, elastomagnetic sensor, ferromagnetic material, finite element method**

I. INTRODUCTION

The article is focused on verification of output signal (voltage) of EMS pressure force. In order to do numerical calculation it is needed to determine the magnetic induction in sensor´s core when the pressure force is and is not affecting it. Simplified EMS model was created to solve the magnetic field in sensor´s core. Magnetic field analysis was realised using the method of finite elements in COSMOS/EMS software for specified frequency and current.

II. ELASTOMAGNETIC SENSOR

A. Elastomagnetic sensor of pressure force 120 kN

Elastomagnetic sensors belong to the group of nonlinear systems. They are operates on the Villari's effect principle, which is based on a change of feromagnetic material permeability proportionally to acting mechanical stress. The permeability increment $\Delta\mu$ proportional to the mechanical stress σ can described by relation [4]

$$\Delta\mu = \frac{2\lambda_{ms}\mu^2}{B_{sef}^2}.\sigma \ ,\tag{1}$$

where λ_{ms} is coefficient magnetostriction for $B = B_s$, B_s is the saturation flux density, μ is magnetic permeability. Equation (1) is valid in the case that the directions of mechanical stress and magnetic field are collinear.
Practical fabrication of elastomagnetic sensor of pressure force 120 kN, which correspond to 100 MPa pressure, is on Fig.1 and geometric computer model created in SolidWorks is on Fig. 2. Windings placement in the core holes is on Fig. 3.

Fig. 1. Elastomagnetic sensors manufactured on KTEEM 120kN
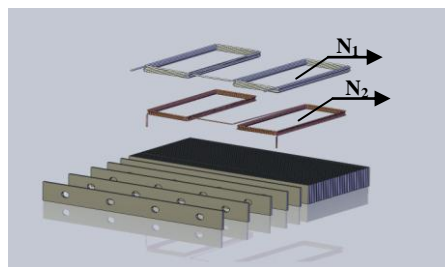
Fig. 2. Design of lamellas and windings
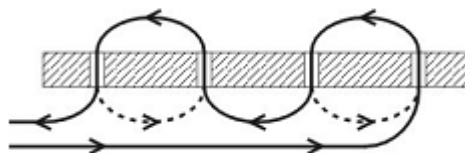
Fig. 3. Placement of windings in the EMS and direction of current flow

B. Relationship between input and output signal of elastomagnetic sensor of pressure force

Input signal for sensor is the external pressure force, output signal is the effective value of voltage, which is induced in secondary winding of sensor and is measured by voltmeter. When defining the voltage on the sensor's output we use the induce law:

▪ In case, that the sensor is unaffected by external pressure force, induced voltage $u_V\ t$ will be:

$$u_V\ t = -N_2\frac{\partial}{\partial t}\left[\int_S \mathbf{B}\ t\ .d\mathbf{S}\right],\tag{2}$$

where N_2 is the number of threads of secondary winding, **B** is the magnetic induction in the sensor´s core unaffected by force in time t, d**S** is normalcy of surface element dS, S is the area of cross-section of sensor´s core.

▪ In case that sensor is affected by external pressure force F induced voltage $u_v^F\ t$ will be:

$$u_v^F\ t\ = -N_2 \frac{\partial}{\partial t}\left[\int_S \mathbf{B}^F\ t\ .d\mathbf{S}\right], \tag{3}$$

where $\mathbf{B}^F$ is magnetic induction in sensor´s core when affected by force in time t.

From metrological point of view it is suitable, that output signal should be related to input signal.

If the magnetostriction coefficient of chosen ferromagnetic material will be positive, output voltage with increasing pressure will begin to fall and $\Delta u_V\ t$ (useful signal) will become negative, therefore we need to adjust the equation:

$$\Delta u_V\ t\ = \left| u_V^F\ t\ - u_V\ t\ \right|, \tag{4}$$

Because magnetic field in the sensor´s core is created by harmonic current $i_1\ t\ = I_{m1}\sin\omega t$ A, where I_{m1} is the maximum value of current $i_1\ t$ and $\omega = 2\pi f$ is the angle frequency of time changes of awaking current. Because of this, magnetic induction and output voltage have time changing behaviour. The question is, which from the values from time changing behaviour shall we measure. Practical measurements showed, that it will be effective value of output voltage .

C   Elastomagnetic sensor´s magnetic field solution

For magnetic potential vector **A** of 3D stationary magnetic field, the following equation is valid:

$$rot\left(\frac{1}{\mu}rot\ \mathbf{A}\right) = \mathbf{J}, \tag{5}$$

where **J** is current density and μ is magnetic permeability of environment in which the field is solved.

For the plane case we assume that the current flows are parallel to the z- axis, it means that

$$\mathbf{J} = \mathbf{k}\ A \tag{6}$$

so only the z component of **A** is present,

$$\mathbf{A} = \mathbf{k}\ A \tag{7}$$

And the equation (5) we can simplify to the scalar elliptic Partial Differential Equation (PDE)

$$-div\left(\frac{1}{\mu}grad\ A\right) = J, \tag{8}$$

where $J = J\ x, y$. $\tag{9}$

For the plane case, the magnetic flux density **B** can be computed as

$$\mathbf{B} = \mathbf{i}\frac{\partial A}{\partial y} + \mathbf{j}\left(-\frac{\partial A}{\partial x}\right) \tag{10}$$

And the magnetic field intensity **H** is

$$\mathbf{H} = \frac{1}{\mu}\mathbf{B} \tag{11}$$

Definition of areas

Area Ω consist of 3 sub-areas (Fig.4),
▪ Sub-area Ω_1, cross-section of wires of primary winding
▪ Sub-area Ω_2, air gap between wire windings and lamella
▪ Sub-area Ω_3, one quarter of lamella's space from ferromagnetic material (possible because of lamella's symmetry)
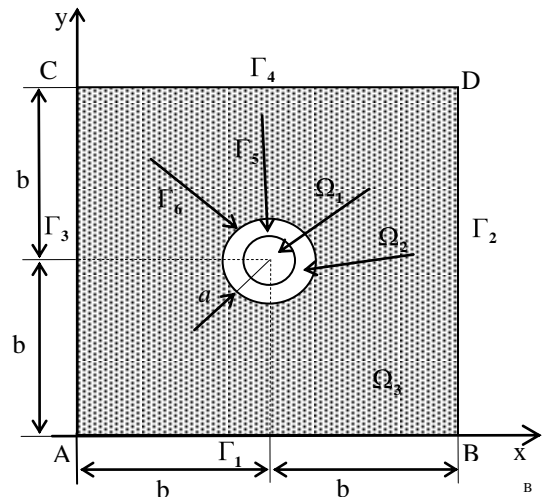
Fig. 4. Definition area of the mathematical model of elastomagnetic sensor

Partial Differential Equation coefficients

For determination magnetic vector potential is used the equation (14) which has in each subdomain the following forms:
▪ in the subdominant Ω_1 :

$$\frac{\partial^2 A_1}{\partial x^2} + \frac{\partial^2 A_1}{\partial y^2} = -\mu_1 J_1, \tag{12}$$

where $\mu_1 = \mu_0$ and $J_1 = 1{,}039.10^7$ Am^{-2}

in the subdominant Ω_2 :

$$\frac{\partial^2 A_2}{\partial x^2} + \frac{\partial^2 A_2}{\partial y^2} = -\mu_2 J_2 \ , \tag{13}$$

where $\mu_2 = \mu_0$ and $J_2 = 0$ A/m^2

- in the subdominant Ω_3 :

$$-\left[\frac{\partial}{\partial x}\left(\frac{1}{\mu_3}\frac{\partial A_3}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{\mu_3}\frac{\partial A_3}{\partial y}\right)\right] = J_3 \tag{14}$$

and $\mu_3 = \mu_3 \ B$, $J_3 = 0$ A/m^2, so for the vectorial magnetic potential we must use non-linear differential equation.

Boundary conditions

The boundary conditions are:
- on the boundaries Γ_1 , Γ_2 , Γ_3 , Γ_4 ,

$$A_3 = 0 \tag{15}$$

- on the boundary Γ_5 between subdomain Ω_3 and Ω_2 :

$$\frac{1}{\mu_3}\frac{\partial A_3}{\partial n} = \frac{1}{\mu_2}\frac{\partial A_2}{\partial n} \ , \tag{16}$$

- on the boundary Γ_6 between subdomain Ω_2 and Ω_1 :

$$\frac{\partial A_2}{\partial n} = \frac{\partial A_1}{\partial n} \ . \tag{17}$$

For magnetic field solution the simplified 3D model from sensor´s core and primary winding of the coils was created. The task was solved as a magnetostatic problem for chosen value of current in nonlinear environment in the cases when sensor is and is not affected by external force. Analysis of magnetic field was realised using the method of finite elements in COSMOS/EMS program. The following graphical interpretations of progresses and graphs of parameters of magnetic field were gained for current density: $J = 1{,}0393772.10^7$ Am^{-2} in wires of primary winding and frequency of current f = 400 Hz (Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9).

Fig. 5. Distribution of magnetic flux density vector in the dependence on pressure force applied to the sensor, F = 0kN
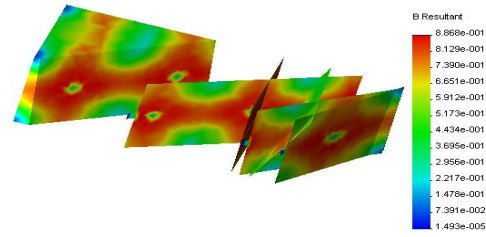
Fig. 6. Distribution of magnetic flux density in the dependence on pressure force applied to the sensor, F = 0kN
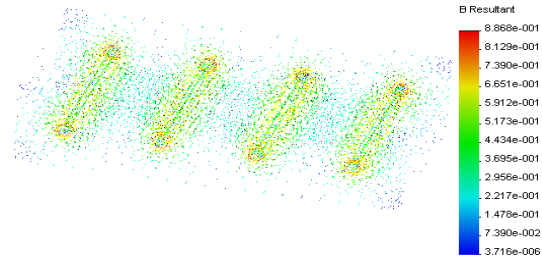
Fig. 7. Distribution of magnetic flux density vector in the dependence on pressure force applied to the sensor, F =120kN
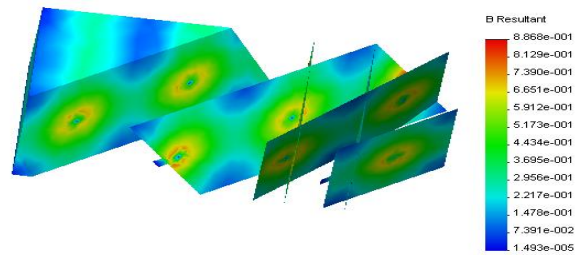
Fig. 8. Distribution of magnetic flux density in the dependence on pressure force applied to the sensor, F = 120kN
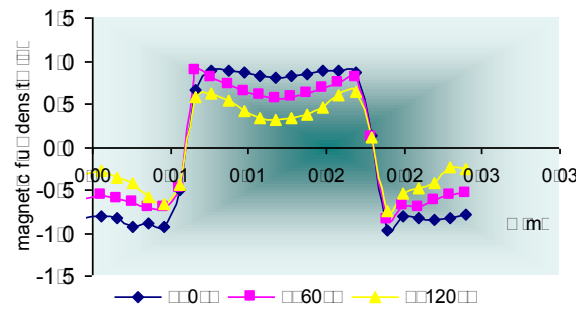
Fig. 9. Distribution of magnetic flux density in the dependence on pressure force applied to the sensor
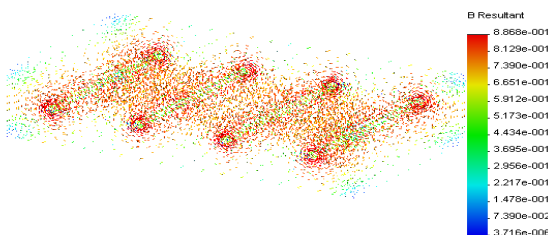
D Measurement and calculation of the output signal EMS

For measuring the effective value of output signal the measuring system was designed and created. Its scheme is pictured on Fig. 10.
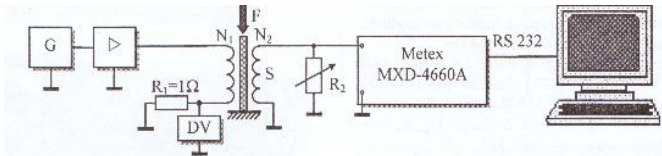
Fig. 10 Apparatus for the measurement of metrological characteristics

When expressing effective value of sensor´s output power it will be convenient to consider that, that behaviour of sensor´s magnetic induction is harmonic too.

Effective values of output voltage and addition of output voltage, directly proportional to pressure force were calculated using the equations (18) and (19) :

$$U_{Vef} = \sqrt{\frac{1}{T}\int_0^T u_V^2 dt} \quad , \tag{18}$$

$$\Delta U_{Vef} = \sqrt{\frac{1}{T}\int_0^T \Delta u_V^2 dt} = \sqrt{\frac{1}{T}\int_0^T |u_V^F(t) - u_V(t)|^2 dt} \quad . \tag{19}$$

These values were compared with the measured values. Both of the values are presented in graphical (Fig.11, Fig.12) and table (Tab.1, Tab.2) form.

F [kN]	Uv [mV] Measured	Uv [mV] Calculated
0	1620	1609
20	1540	1523
40	1435	1428
60	1304	1324
80	1172	1171
100	1046	1036
120	929	928

Tab. 1. Calculated and measured values of output sensor voltage

F [kN]	ΔUv [mV] Measured	ΔUv [mV] Calculated
20	80	86
40	185	181
60	316	285
80	448	438
100	574	573
120	691	681

Tab. 2. Calculated and measured values addition of output sensor voltage
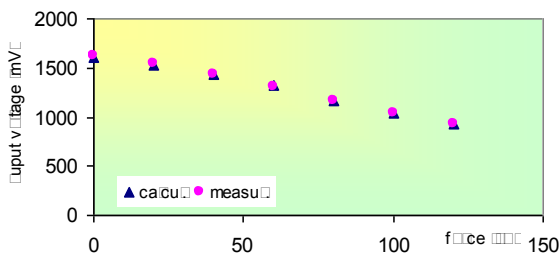
Fig. 11. Calculated and measured values addition of output sensor voltage
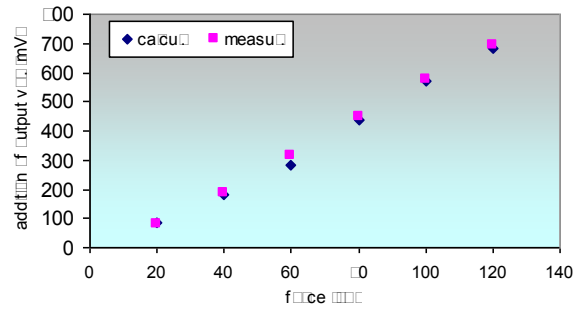
Fig. 12. Calculated and measured values addition of output sensor voltage

II. Conclusion

In the article the relation between input and output EMS pressure force signal was verified. The values of output voltage are dependent on magnetic induction. With the increased pressure force the magnetic induction in sensor´s core decreases, as well as output voltage of the sensor. This was verified with practical measurements.

The calculated and measured results comparison shows the differences. The differences are most probably caused by using simplified assumption.

Acknowledgment

References

[1] I. Tomčíková, Modelling of magnetic field of elastomagnetic sensor of pressure force (in Slovak). In: Habilitate thesis, Košice, 2009, 83 pages.
[2] I. Tomčíková, D. Špaldonová, Elastomagnetic Sensor Field Determinati Using Matlab, Acta Electrotechnica et Informatica,Vol.. 7,No. 3, 2007, pp. 74-77, ISSN 1335-8243.
[3] L. Haňka, Electromagnetic Theory. SNTL, Praha, 1982
[4] K. Zehnula, Sensors of Non-electric Quantities. SNTL, Prague, 1977
[5] D.Mayer, B. Ulrych, M.Škopek, Solution of electromagnetic fields using modern software products, Journal EE, Vol. 7, 2001, No.1., No.2.
[6] J. Vojtko, et al., Utilization of Elastomagnetic effect in Force and torque measurements. Proceedings of the 2nd Conference, MWT TU FEI Košice, 2004.
[7] A. Hodulíková, Measurement of Force Using Elastomagnetic Effect, 5th PhD.Student Conference, TU FEI Košice, 2005.
[8] I. Tomčíková, D. Kováč , I. Kováčová, Stress field distribution in magnetoelastic pressure force sensor. In : Communications, no.1, 2010, pp. 16-19.

14

Electromagnetic engraving tool and its application

[1]*Mišél BATMEND,* [2]*František BANÍK*

[1]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic
[2]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]misel.batmend@tuke.sk, [2]frantisek@banik.sk

Abstract—**The proposed paper describes the principles of construction of an electromagnetic diamond percussion tool used for stone engraving. The tool foundation is an electrically driven pointed diamond engraving by hammering on solid, polished surfaces. The paper includes measurements of tool prototype, describing its basic features and potential applications. Finally, the fundamental image processing methods, necessary for achieving satisfactory results using said tool, are introduced. All the methods are experimentally examined.**

Keywords—**percussion tool, headstone engraving, photo etching**

I. INTRODUCTION

Nowadays, a majority of headstone portraits is handmade, although machine engraving becomes trendy. The machines are XY CNC plotters using either laser tool [1], or milling cutter [2]. The proposed tool demonstrates another way of engraving, relating mostly to methods and tools used for handmade portraits. The handheld tool is usually machine of the type in which alternating electric current is used to bring about the vibratory movement of sharply pointed cutter used for engraving (see [3]). Control of single stroke of the cutter is impossible due to the method of driving of actuating solenoid. This would be useless in terms of handheld tools, but when attempting to use such tool connected to CNC plotter, it turns out to be crucial. Controlling of single strokes allows the plotter to build up a mosaic of dots representing picture or portrait based on black & white model image.

When driving the solenoid by buck-boost converter, the duration of striking and recovering of tool to its original position can be easily changed. Also the magnitude of current fed to the solenoid can be adjusted. Experimenting with these parameters, rather optimal functionality of the tool can be obtained.

II. ENGRAVING TOOL DESIGN

A. Mechanical properties of the tool

The tool (Fig. 1) compounds of plunger core actuable by solenoid to conduct the striking operation and a set of spring members to recover the plunger core back to the original position. The lower end of plunger core is equipped with pointed diamond.
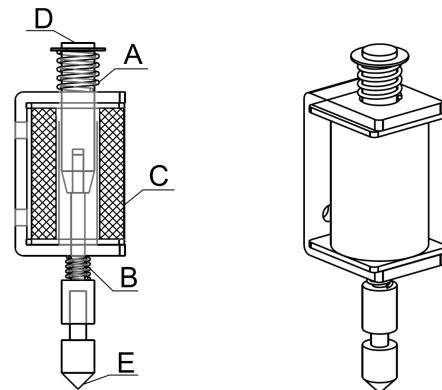
Fig. 1 Percussion tool: A – upper spring, B – lower spring, C – solenoid, D – plunger core, E – pointed diamond

The A spring recovers the plunger core after striking operation, the B spring serves for dumping the back stroke of core while recovering. Moreover, when properly timed, it raises the energy of following stroke. The spring rates are related as follows:

$$k_A > k_B \qquad (1)$$

B. Dynamic properties of the tool

Feeding the electric current into the solenoid upraises a magnetic field which causes the ferrite plunger core to slide into the hole in center of solenoid. Core together with diamond is forced to make a vertical movement. If a polished stone desk is put into specific distance, the diamond will hit the desk which will cause disruption of polished surface. This disruption usually appears as a brighter dot on a darker stone. In this state, the plunger core is in its lowest position, whilst the diamond is touching the desk. The A spring is pressed (see Fig.2 time t=0,1s).

Interrupting the current flow in solenoid will cause the magnetic field to disappear. The A spring will recover the plunger core to its original (or zero) position. The moment of inertia of core together with set of springs A and B will cause the overshoot and oscillation around the original position. Finally the core will reach its steady state, when the forces of both springs equalize. Fig. 2 depicts the plot of position of diamond tip performing one stroke.
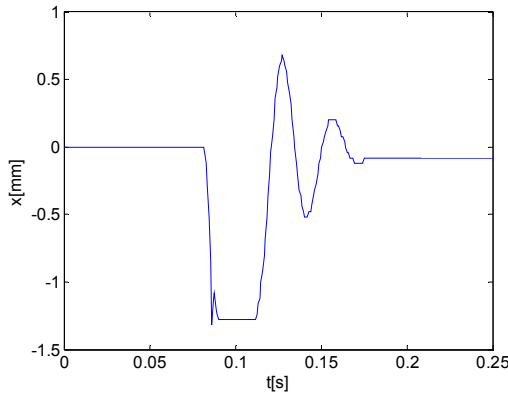
Fig. 2 The diamond tip position during one stroke

The plot was captured by measuring on prototype tool, constructed according to Fig. 1. As a position measuring device, a linear quadrature sensor detecting 25 impulses/mm was used. The time base sampling period was equal to 0,819 ms. The sensor was attached to the plunger core in the way that its influence (additional mass) on measurement was minimized.

As can be seen in a plot, after hitting the desk for the first time, the diamond tip will bounce off and hit the desk again. To disrupt the surface of the desk, energy of the first hit is decisive. Concerning the maximal frequency of strokes, it is sufficient to set the timing, so that the tip will hit the desk only once. The energy of bounce can be then used to help recover the plunger core into zero position (where x=0). With appropriate timing, the energy of overshooting core can be also added to following stroke.

III. DYNAMIC ANALYSIS AND EXPERIMENTAL RESULTS

To identify the tool maximal percussion frequency and its relation to magnitude of actuating force, set of measurements has been taken. Stroke frequency is relevant parameter, because it has significant influence on overall speed of engraving. One could presuppose that stronger force would allow the core to oscillate faster and vice versa. Adjusting of actuating force is feasible due to a buck – boost converter which is driving the solenoid. Such type of converter is suitable for this application because it allows faster demagnetization of coil compared to simple buck converter.

The detailed schematic of converter is shown in Fig. 3. It includes pair of MOSFET transistors and diodes connected to solenoid represented by inductor and serial resistor. Transistors are driven by high and low side driver with bootstrap circuit. Since the converter is supplied by single voltage source (24V), secondary voltages necessary for driver and optocoupler operation (15V and 5.1V respectively) are derived using Zener diodes. The converter is driven by external pulse generator which is decoupled by pair of high-speed optocouplers.

The electromagnetic actuating force is proportional to current flowing trough solenoid. The current magnitude is set by modulating the solenoid voltage by PWM with carrier frequency of 10 kHz. Time of feeding and non-feeding of solenoid can be set as well. Adjusting these parameters, following measurements have been taken.

Fig. 4 shows a measurement with voltage duty cycle d=0.5, and switching times t_{on}=8ms and t_{off}=9ms. The first two strokes have longer switching times (t_{on}=30ms, t_{off}=20ms) in order to define a full swing of diamond tip.

Shortening of switching times, while using the same voltage duty cycle leads to strike irregularities (see Fig. 5). The tip does not reach the zero position before a following strike is actuated.

Using a voltage duty cycle d=0.5 yields the maximum percussion frequency:

$$f_{d0.5} = \frac{1}{t_{on}+t_{off}} = \frac{1}{8\times10^{-3}+9\times10^{-3}} = 58.8 Hz \qquad (2)$$

Increasing the voltage duty cycle leads to boosting the current flow through solenoid which concludes in rise of magnetic force actuating the plunger core. Thus the stroke on the stone desk is stronger. Therefore a diameter and depth of a bright dot increases. Fig. 6 depicts a measurement with voltage duty cycle d=0.7.

When using the t_{on} = 8ms and t_{off} = 9ms, bouncing of the tool tip can be noticed (see Fig. 6). To avoid this effect, shortening of the on time to t_{on}=7ms is necessary (see Fig. 7).
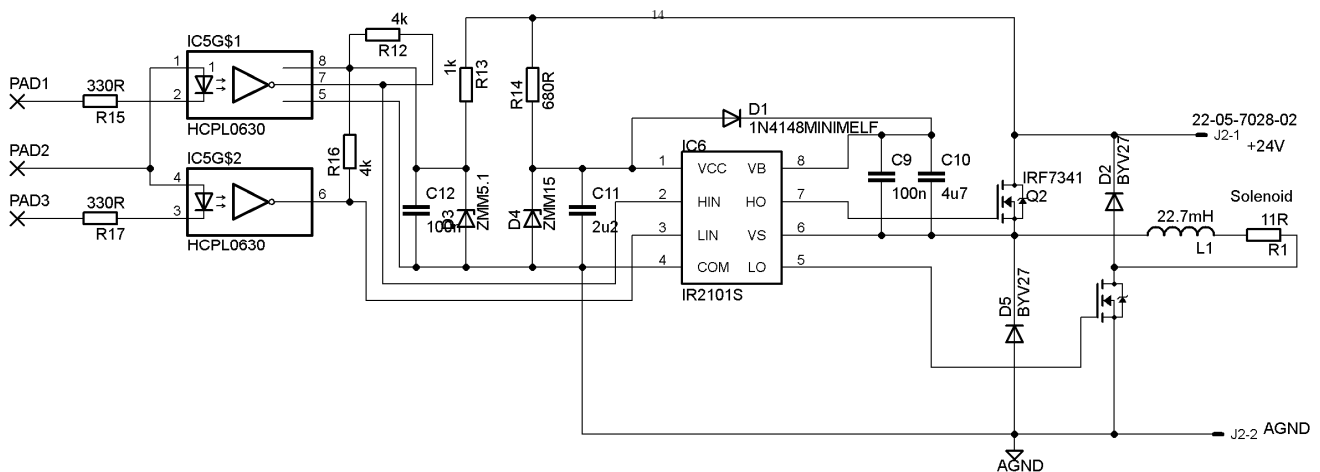
Fig. 3 Buck – boost converter, complete schematic
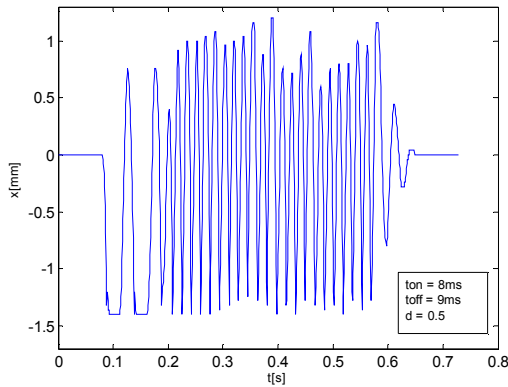
Fig. 4 The diamond tip position. Strokes with duty cycle of 0.5. Maximal frequency, no overlapping of strokes.
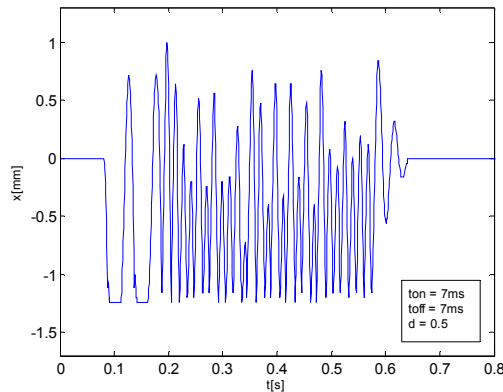
Fig. 5 The diamond tip position. Strokes with duty cycle of 0.5. Higher frequency causes irregular strokes. The plunger doesn't reach its zero position before following stroke.
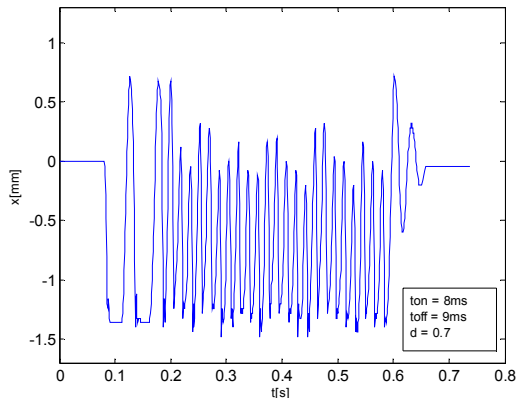
Fig. 6 The diamond tip position. Strokes with duty cycle of 0.7. Lower frequency causes undesirable bouncing of diamond tip.
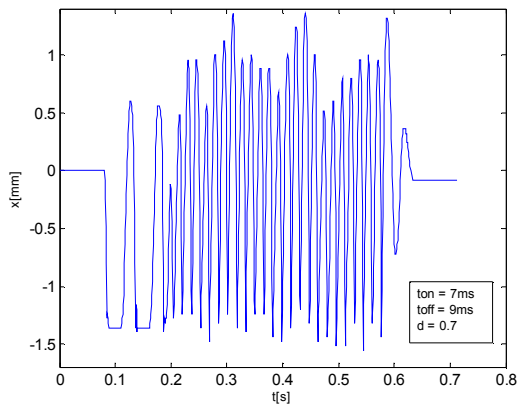
Fig. 7 The diamond tip position. Strokes with duty cycle of 0.7. Bouncing can be prevented by increasing strike frequency.

Decreasing of voltage duty cycle leads to a softer stroke, which will show as smaller bright dot on the stone desk. When using parameters d = 0.3, t_{on} = 8ms and t_{off} = 11ms, the tool tip does not even reach the desk (Fig.8), because the on time is too short to let the tip get to the desk. Rising the on and off times solves the problem (Fig. 9). Hence, the strike frequency decreases:

$$f_{d0.3} = \frac{1}{30 \times 10^{-3}} = 33.33 Hz \qquad (3)$$
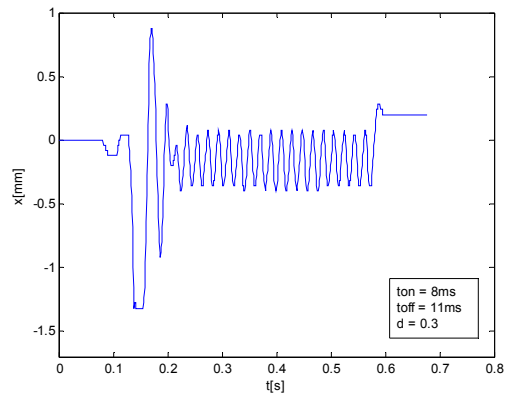
Fig. 8 The diamond tip position. Strokes with duty cycle of 0.3. The actuating force is too small to let the tool tip get all the way to the stone desk.
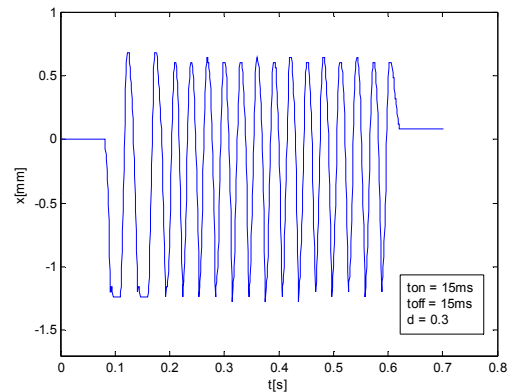
Fig. 9 The diamond tip position. Strokes with duty cycle of 0.3. Decreasing of frequency leads to full swing of the tool tip,

To test the accuracy of set on and off times corresponding to particular voltage duty cycles, a simple image based on a model image (Fig.10 a) was engraved. The model image is an 8 bit grayscale. The brightness rises from left to right starting with darkest finishing with brightest. The image was transformed to a binary black & white by conversion algorithm chosen according to [4]. The halftoning algorithm uses cluster dot dithering method, which was chosen from variety of halftoning methods as a most suitable for stone engraving. Subsequently, the image was engraved three times, each time with different voltage duty cycle and corresponding on and off times. After engraving, the image was daubed with a white paint, which settles in deeper dots and thus enhances a contrast between the image and the stone desk.
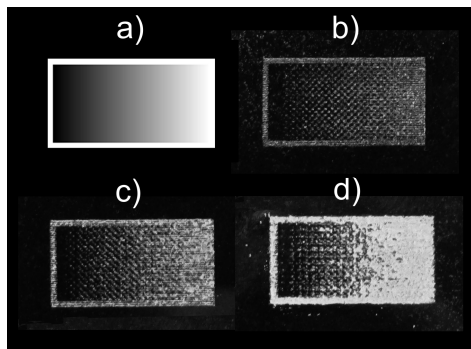
Fig. 10 Image engraved using different voltage duty cycles and on and off times. Scale 1:1.

In Fig. 10 b) a softer stroke, actuated by smaller force (voltage duty cycle d = 0.3) was used. In Fig. 10 c) medium stroke was used (d = 0.5). Finally on Fig. 10 d) stronger stroke was used (d=0.7). Brightness levels of particular images are clearly distinguishable. The darkest image dots are so fine that no paint has settled in. In case (c), bright tones are brighter than in case (b). Transition from dark tones to bright ones is still smooth enough. In case (d), the bright tones are indisputable brightest. On the other hand, the dark-to-bright transition is rough and unsmooth. One can notice some kind of saturation in the bright tones before reaching a maximum brightness. This is caused by overlapping of single dots, since they are too big and create unified piece of white paint even in areas where certain amount of dark dots is expected.

IV. CONCLUSION

By detailed analysis and performed experiments with proposed tool we gained a good concept of application capabilities of the tool for engraving portraits into stone desks. We accomplished following conclusions:

1. Size and depth of engraved dot depend mostly on stroke strength. There is only a small influence of duration of stroke on these properties.

2. Duration of stroke should be considered mostly in terms of reaching the maximum frequency of strokes.

Regarding the experiment with engraved image, one can conclude that contrast of engraved image could be enhanced by using variable stroke actuating force. To get as wide scale of tones as possible, the strongest stroke should be also used. Eliminating of bright tones saturation and smoothing of dark-to-bright transition would be necessary.

All the experiments were performed using constant distance of the tool tip from stone desk (approx. 1.3 mm). Reducing the distance yields increase in stroke frequency with trade-off of stronger stroke for the dot of the same size. Using distance 1.3 mm is adequate also because the stone desk is not perfectly flat.

Considering the other tools for engraving, the proposed tool has simple construction and is also cheap. In contrary to milling cutter, it does not wear. When used as a XY plotter tool, the drawback is lower applicable resolution of engraved image (approx. 180dpi). This resolution constraint is introduced by size of the smallest engraved dot which refers to one image pixel. Laser machine resolution is at least 300 dpi.

REFERENCES

[1] Z12 Monument Etching Laser, available on the internet: http://www.kernlasers.com/product_pages/z12.htm
[2] G10 stone image engraving machine, available on the internet: http://www.acmetek.biz/g10.shtml
[3] H. L. K. Johnson, "Improvements in or relating to Electrically Driven Engraving machines", GB536146, May 1941, Patent specification.
[4] M. Batmend, "An automated headstone photo engraving" (proceedings from conference), 10th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice, I. Edition, pp. 171–173, ISBN 978-80-553-0423-6.
[5] GONZALEZ, R. C. – WINTZ, P.: *Digital image processing*.2nd edition, Addison-Wesley publishing company, 1987, ISBN 0-201-11026-1
[6] Jarvis, J.F., C.N. Judice, and W.H. Ninke, "A Survey of Techniques for the Display of Continuous Tone Pictures on Bi-Level Displays," Computer Graphics and Image Processing, vol. 5, pp. 13-40, 1976.

Experimental Examination of the Shadowing Effect for the Short Range UWB Radar

Daniel URDZÍK, Jana ROVŇÁKOVÁ

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

daniel.urdzik@tuke.sk, jana.rovnakova@tuke.sk

Abstract—**Ultra wideband (UWB) radar appears as the suitable technology for detection and tracking of people in critical situations and environments. However such enviroments are in many cases small areas (e.g. rooms) in which the phenomenon of mutual shadowing between people can occur. In this paper, the theoretical basis of the shadowing effect is presented. To examine mutual shadowing in detail, the measuring campaign with short range UWB radar was carried. Its preliminary results are described and analyzed in this paper. They indicate the approximative shape of the dead zones created by the shadowing person. In the conclusion, the solution of this problem is briefly discussed.**

Keywords—**UWB radar, shadowing effect, shadows, dead zones, detection, tracking.**

I. INTRODUCTION

In recent years the use of short range UWB radar is becoming an attractive solution for the localization purposes. The UWB radar system is the special type of radar and can be used to detect and track moving targets in critical environments with an advantage. One of the features of the UWB radar systems is their use for the applications including through wall and through fire detection and tracking of moving targets during security operations, protection of reservoirs, power plants and other critical infrastructures against to a terrorist attack, detection of trapped people after an avalanche or earthquake, etc. UWB radar emissions applied for these scenarios are at a relatively low frequency (typically between 100 MHz and 5 GHz). Additionally, the fractional bandwidth of the signal emitted by UWB radar is very large. The result is a UWB radar with a fine resolution that also has an ability to penetrate many common materials. Then, such devices are able to detect moving person by measuring changes in the impulse response of the environments [1]. In the above outlined applications, a short range UWB radar (range up to 20-25m) is usually applied. Therefore we can refer these scenarios as short range tracking.

The problem of short range detection and tracking of moving persons have been studied e.g. in [2], [3], [4], [5], [6]. However, the problem of multiple human tracking in real complex environment has been less well addressed. Our experiments from different measurement campaigns with the short range experimental UWB radar, which were dedicated to detection and tracking of multiple moving persons showed that the detection of multiple persons is met with difficulties due to the phenomenon of shadowing.

The effect of shadowing have been described e.g. in [7], [8], [9], and [10] where the shadow is defined as the absence of radar illumination because of the presence of stationary objects e.g. metal boards within the illuminating radar beam. The shadows behind the obstacles are caused by reflection of emitted radar signals from the obstacles. However in the case of indoor measurement with UWB radar another type of shadowing may occur. This type of shadowing is present when the person is between the radar and the strong stationary electromagnetic scatterer (i.e. the scatterer is in the person's shadow). If no target is present, the scatterer as the stationary object appears as the background to the radar and it is subtracted from all incoming raw radar signals. On the other hand, if the person starts moving in front of the scatterer, the reflection from it are not received by the radar but the background subtraction algorithm still remembers them and continues in their subtraction. This results in appearing of strong reflections from the static scatterer in the radar signals with subtracted background exactly in the observation time instants during which the person shadowed it.

The third type of shadowing in through-wall radar applications occurs in the case when more people are moving in a small area (e.g. room). If in such scenarios one or more persons are close to the transmitting or receiving antennas and the rest of them are far from the antennas, we found out that the effect of mutual person shadowing will take place. This phenomenon can be theoretically explained by the fact that due to the frequency band employed by the UWB radar a person moving nearby the radar antennas reflects and absorbs the energy of electromagnetic waves emitted by the transmitting antenna but only a negligible part of electromagnetic wave energy is transmitted through the person body to a certain region located behind the person where the second target can be located. The similar effect can be identified, if the person moving nearby the radar antennas is located between the second target and receiving antenna which results in that this person absorbs the energy of electromagnetic waves reflected by the second target and hence, only a negligible part of energy of electromagnetic waves reflected by the second target can be received by the radar. The persons which create the shadowing effect will be denoted as shadowing persons hereinafter. The presence of such shadowing persons which absorbs large portions of energy creates the dead zones where radar practically cannot detect and track any of the persons which are located within these zones.

According to our best knowledge, this type of mutual shadowing effect has not been described and experimentally confirmed for the purposes of detection and tracking of moving people. The intention of this paper is to describe the exper-

(a)   (b)   (c)

Fig. 1. The setup of UWB radar device in the school dining room for measurement of the shadowing effect, (a) view from behind of the radar device, (b) view from behind of the area of interest, (c) the scheme of the area with the reference points.

iments which were dedicated to the problem of shadowing effect and to confirm that one short range UWB radar is usually not sufficient for detection and tracking of multiple targets moving in the small area.

II. MEASURING CAMPAIGN

To qualitatively describe the effect of shadowing the measurement campaign with specific scenarios had to be managed. The measurements took place in the large room (school dining room) without any additional obstacles to analyze this phenomenon (Fig. 1). The scenarios were chosen in such a way to examine the consequencies of the shadowing person presence on detectability and trackability of the other people in small area of interest.

There were five basic scenarios measured. Each scenario consisted of varying number of the targets moving along different predefined trajectories. We performed also additional experiments with antenna polarization. The goal of these experiments was to find out what effect has the polarization of emitted signal on the quality of detection and tracking by the UWB radar device. Concluding part of the measurement campaign was realization of the measurements with different vertical positions of the antennas. Each of the basic scenarios was measured then with and without the shadowing person utilizing vertical or horizontal polarization of antennas located in the height of 15 or 75 cm. The shadowing person was standing gradually in front of each antenna (i.e. three cases for each basic scenario). The shape of the trajectories was specific to analyze which parts of the trajectory the radar will not be able to track and hence how large is the dead zone behind the shadowing person.

The raw radar signals were acquired by the UWB pseudo-noise radar using maximum-length-binary-sequence (M-sequence) as the stimulus signal [11], [12]. The system clock frequency of the experimental M-sequence UWB radar is about 4.5 GHz, which results in the operational bandwidth of about DC - 2.25 GHz. The M-sequence order emitted by radar is 9, i.e. the impulse response covers 511 samples regularly spread over 114 ns. This corresponds to an observation window of 114 ns leading to an unambiguous range of about 16 m. The radar device was equipped with one transmitting antenna Tx and two receiving antennas Rx_1, Rx_2 (Fig. 1(a)).

III. PRELIMINARY EXPERIMENTAL RESULTS

To illustrate the mutual shadowing effect between people moving in a small area on real radar signals, the following experimental measurement was chosen. One person was moving in front of the radar system (Fig. 1(a)) along the trajectory of shape "S" through the reference points P1-P2-P3-P4-P5-P6-P7-P8-P9 (Fig. 1(c)). The second person was standing in the distance of 1 m in front of one of three antennas. In the next the scenario, when the shadowing person was located in front of transmitting antenna, is labeled as Scenario Tx and in the case of person standing in front of one of the receiving antennas, it is reffered as Scenario Rx1 and Scenario Rx2. All three scenarios are compared to the basic scenario with trajectory "S" without presence of shadowing person.

The raw radar signals have been processed by the complex signal processing procedure for multiple target scenarios designed for short range UWB radar with one transmitting and two receiving antennas proposed in [13]. It consists of seven phases, namely background subtraction, weak signal enhancement, detection, time of arrival (TOA) estimation, wall effect compensation, localization and target tracking. For the purposes of our experiment, the phase of weak signal enhancement was omitted to analyze the whole undistored impact of shadowing effect. As the measurement was done in the free area, the wall effect compesation phase was omitted, too.

The processing results of the basic scenario are shown in Fig. 2(a)-2(c). The first two pictures have the form of radargram - a two dimensional picture, in which the vertical axis is related to the time propagation of the radar signal (impulse response) and the horizontal axis is related to the observation time. These radargrams represent the detector output for antena Rx1 (Fig. 2(a)) and Rx2 (Fig. 2(b)). As we can see from there, the radargrams are quite similar, only finely time shifted, what results from small and symetrical distance of the receiving antennas from the transmitting antenna (Fig. 1, Tx in the middle between Rx1 in the right and Rx2 in the left). The detector output consists of binary data, where values "1" (black points in Fig. 2(a)-2(b)) represent reflections from moving targets. They create so called primary target trace. In Fig. 2(a)-2(b), two such traces can be visualy recognised. The first one, starting at propagation time instant of 20 ns
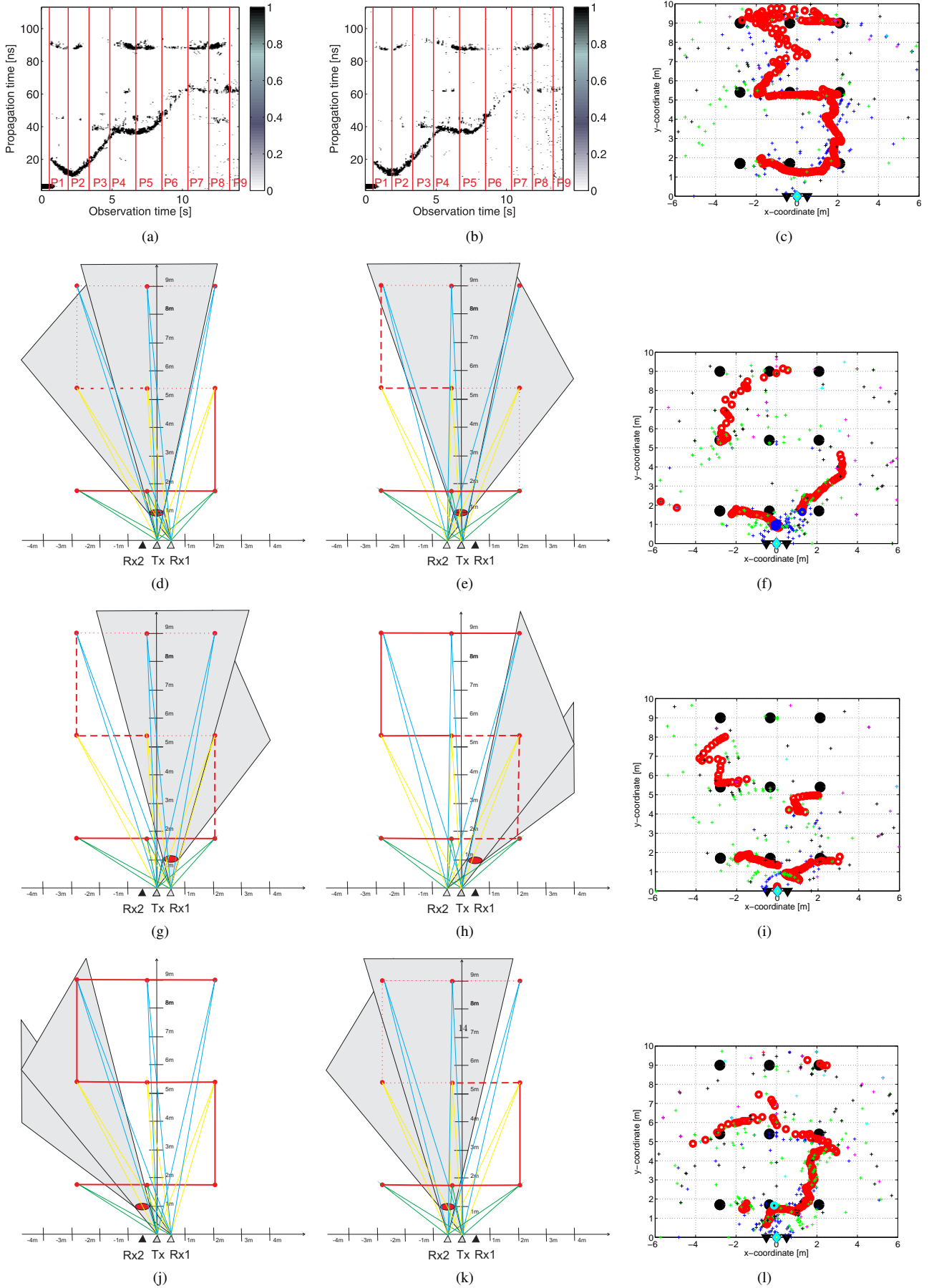
Fig. 2. Preliminary experimental results. Basic scenario: (a) detector output from Rx1, (b) detector output from Rx2, (c) estimated target track; Scenario Tx: (d) spatial representation of detector output from Rx1, (e) spatial representation of detector output from Rx2, (f) estimated target track; Scenario Rx1: (g) spatial representation of detector output from Rx1, (h) spatial representation of detector output from Rx2, (i) estimated target track; Scenario Rx2: (j) spatial representation of detector output from Rx1, (k) spatial representation of detector output from Rx2, (l) estimated target track.

and finishing roughly at 65 ns, corresponds with reflections from the person walking along trajectory "S". The second discontinuous trace (around 90 ns) illustrates the shadowing of a large static object by the moving person. In our case it was the glass wall with furniture located around 13 m from the radar system (Fig. 1). The final target track in Fig. 2(c) was obtained by processing of the detector output by the methods of TOA estimation, localization and tracking. It can be seen from this picture that the final estimated track corresponds quite well with the true target trajectory.

To more demonstratively evaluate the mutual shadowing effect in scenarios with shadowing person, the following transformation of radargrams from time domain to spatial domain has been done. The detector outputs have been divided according the reference points P1-P9 to the time intervals, in which the person was moving from the point Pi to point P(i+1) for i=1,2,...,8 (Fig. 2(a)-2(b)). In each time interval, the quality of received reflections from moving target was scored according the scale: strong reflections (red solid line in spatial domain), weak reflections (red dashed line) and none reflections (red dotted line). The pictures created in such a way were complemented by the oval representation of shadowing person with diameter of 50 cm. This enabled to outline the expected shadow (dead zone) by utilization of the geometrical optics methods.

The results for Scenario Tx are depicted in Fig. 2(d)-2(f). As can be seen from the first two pictures, both antennas received strong reflections from the target only when it was moving in the vicinity or radar antenna system (positions P1-P3 and partially P4). In the rest of trajectory, the received reflections were weak or none. It resulted in incomplete estimation of target track (Fig. 2(f)). Shadows outlined in Fig. 2(d)-2(e) correspond well with measured radar data excepting the interval between positions P5 and P6 scanned by the antenna Rx1. Here we expected none reflections instead of received weak.

The Scenario Rx1 is analyzed in Fig. 2(g)-2(i). In this case, the reflections received by the antenna Rx1 are markedly shadowed by the standing person (Fig. 2(g)). On the other hand, the antenna Rx2 "has seen" almost whole trajectory. As the target locations are calculated based on combination of data from both receiving antennas, the resulted track is again very bad estimated (Fig. 2(i)). The outlined shadows correspond very well with measured radar data.

The last case is Scenario Rx2 demonstrated in Fig. 2(j)-2(l). Here the output from the antenna Rx1 was surprising because Fig. 2(j) should be similar to Fig. 2(h). Instead of this, the primary target trace was even stronger as in the basic scenario without shadowing person. Fig 2(k) matched with expectations. Almost whole trajectory was shadowed what resulted in estimation only of a torso of the true target track (Fig. 2(l)).

IV. CONCLUSION

The preliminary experimental results obtained by the processing of real UWB radar signals have confirmed presence of mutual shadowing effect between people moving in small area. It was shown that a person situated in front of any of the antennas markedly corrupts the final track estimations. The shape of dead zones (shadows) was approximated based on the methods of geometrical optics. In the most of analyzed cases,

this approximation has corresponded with the received radar data. However the dead zone properties need to be examined from different aspects. This is the main aim in the processing of remaining scenarios of the described measuring campaign.

The solution of the outlined shadowing phenomenon seems to be in the utilization of the UWB sensor network consisting at least from two sensors located in different positions. Our first results in this area indicate a perspective way for detection and tracking of multiple targets (people) moving behind walls.

REFERENCES

[1] P. Withington, H. Fluhler, and S. Nag, "Enhancing homeland security with advanced UWB sensors," *Microwave Magazine, IEEE,* vol. 4, no. 3, pp. 51–58, Sep. 2003.

[2] S. Chang, R. Sharan, M. Wolf, N. Mitsumoto, and J. W. Burdick, "People Tracking with UWB Radar Using a Multiple-Hypothesis Tracking of Clusters (MHTC) Method," in *Int J Soc Robot (2010) 2.* Springer, Jan. 2010, p. 3 18.

[3] S. Z. Gurbuz, W. L. Melvin, and D. B. Williams, "Comparison of radar-based human detection techniques," in *Proc. 41st Asilomar Conf. on Signals, Systems and Computers*, 2007, pp. 2199–2203.

[4] S. S. Gauthier and W. Chamma, "Surveillance through concrete walls," in *Proc. of SPIE -C3I Technologies for Homeland Security and Homeland Defense III*, Sep. 2004, pp. 597–608.

[5] S. Nag, M. A. Barnes, T. Payment, and G. Holladay, "Ultrawideband through-wall radar for detecting the motion of people in real time," *Proc. SPIE-Radar Sensor Technology and Data Visualization*, vol. 4744, pp. 48–57, Jul. 2002.

[6] S. S. Ram and H. Ling, "Through-Wall Tracking of Human Movers Using Joint Doppler and Array Processing," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 537–541, Jul. 2008.

[7] A. Theil and L. van Ewijk, "Radar performance degradation due to the presence of wind turbines." Radar Conference, IEEE, 17-20 April 2007, pp. 75–80.

[8] R. Bindlish and A. P. Barros, "Parameterization of vegetation backscatter in radar-based, soil moisture estimation," *Remote Sensing of Environment*, vol. 76, no. 1, pp. 130–137, April 2001.

[9] S. Fujita, T. Sakamoto, and T. Sato, "An accurate UWB radar imaging method using indoor multipath echoes for targets in shadow regions." Indoor Positioning and Indoor Navigation (IPIN), Sep. 2010, pp. 1–7.

[10] L. Vinagre and K. Woodbridge, "Secondary surveillance radar monopulse target azimuth error estimation due to obstacle shadowing." Radar Conference, IEEE, Apr. 1999, pp. 166–171.

[11] D. Daniels, *Ground Penetrating Radar, 2nd ed.* London, U.K.: The Institution of Electrical Engineers, Jul. 2004, ch. M-sequence radar.

[12] J. Sachs, R. Herrmann, M. Kmec, and P. Peyerl, "Modified M-sequence UWB-radar," in *German Microwave Conference (GeMiC)*, Karlsruhe, Germany, March 2006.

[13] J. Rovňaková, *Complete signal processing for through wall tracking of moving targets.* Germany: LAP LAMBERT Academic Publishing, Sep. 2010.

14

Finding of robots direction in space using harmonic potential field

[1]*Michal KAĽAVSKÝ,* [2]*Miroslav ŤAHLA*

[1]Dept. of Electrical Drivers and Mechatronics, FEI TU of Košice, Slovak Republic
[2]Dept. of Electrical Drivers and Mechatronics, FEI TU of Košice, Slovak Republic

[1]michal.kalavsky@tuke.sk, [2]miroslav.tahla@tuke.sk

Abstract — **Harmonic potential field is one of the most using method for planning path of robots in space with obstacles. This method is mostly using for static space, what means that is known size of area, position of goal and position of obstacles. Method harmonic potential field belongs to methods of planning on grid.**

Keywords—**potential, Dirichlet problem**

I. Introduction

In recent years, we know several methods that are used for finding of direction of robots in space. To this methods belong to for example Bug algorithms, exact planning, planning on grid and neuron networks. Using of particular method depends on criteria that chosen method must satisfy these criteria at solution of mission. In this paper is described in detail method harmonic potential field which belongs to planning on grid.

II. Method Harmonic Potential Field

One of the most used method for planning path of robots in space is method potential field. As in introduction was considered, this method belongs to planning on grid. What exactly means planning on grid? Imagine the space where robot is moving. This space is divided on cells where every cell represents a certain part of this space and every cell has a certain evaluation. By this process is achieved that space is allocated on certain number of cells or sections what introduces the grid with a certain number of cells.

A. Potential fields

Potential fields are based on construction of virtual maps that are simulated a real space. This means that by potential levels is created virtual map of real space and this robot has at command for finding direction in space. Robot is moving by direction on the top what means that is moving in direction the highest of growth of potential.

As we can see from name potential field, at this method is used potential. By potential create potential levels that are helpful at finding of robot direction that robot will arrived to the goal position. Cell that is represented start position has the lowest evaluation on grid. Cell that is represented goal position has the highest evaluation on grid. By this evaluation is reached that robot is moving on the top. In space can be obstacles so robot must be capable to avoid these obstacles. Therefore must be executed one condition that all obstacles must be convex. If this condition is not executed then robot can be stuck in local minimum. This means that robot arrives to the position that is not goal position and then is embedded in cycle. Local minimum is the basic weakness method potential fields. However, this problem is possible to solve using method harmonic potential fields where exists only one minimum, global minimum represented by goal position. [2],[3]

Consider that method of potential fields is based on potential and potential levels, is needed define what is potential. Consider electrical potential. By analogy we can define method potential fields.

Electrical potential is scalar physical quantity that is described potential energy of electrical unit charge in constant electrical field. Thence we can define potential between two points as quantity of work needed for transplantation of this electrical unit charge from one point to other.

In accordance to [9] is apparently from definition that

$$d\varphi = -\vec{E}.$$

$$(1)$$

where $\vec{E}$ is vector of electrical field intensity and $\overrightarrow{dr}$ is orbital element.

For

$$\vec{E} = Ex\vec{i} + Ey\vec{j} + Ez\vec{k}$$

and for

$$d\vec{r} = dx\vec{i} + dy\vec{j} + dz\vec{k}$$

dot product has form:

$$\vec{E}\,d\vec{r} = Ex\,dx + Ey\,dy + Ez\,dz$$

$$(2)$$

For $\varphi=\varphi(x,y,z)$ then we can express total differential $d\varphi$ shaped:

$$d\varphi = \frac{\partial \varphi}{\partial x}dx + \frac{\partial \varphi}{\partial y}dy + \frac{\partial \varphi}{\partial z}dz$$

$$(3)$$

Equation (1) is possible re-write by equations (2) and (3) to the form:

$$\frac{\partial \varphi}{\partial x}dx + \frac{\partial \varphi}{\partial y}dy + \frac{\partial \varphi}{\partial z}dz = -Ex\,dx - Ey\,dy - Ez\,dz$$

(4)

Thence on the basic of independence of parts x, y, z result parts of vector $\vec{E}$ in form:

$$Ex = -\frac{\partial \varphi}{\partial x} \qquad Ey = -\frac{\partial \varphi}{\partial y} \qquad Ez = -\frac{\partial \varphi}{\partial z}$$

(5)

Vector of electrical field intensity $\vec{E}$ is possible express shaped:

$$\vec{E} = Ex\vec{i} + Ey\vec{j} + Ez\vec{k} = -\frac{\partial \varphi}{\partial x}\vec{i} - \frac{\partial \varphi}{\partial y}\vec{j} - \frac{\partial \varphi}{\partial z}\vec{k}$$

(6)

Using operator *grad* in Cartesian coordinate system we receive expression

$$grad\ \varphi = \frac{\partial \varphi}{\partial x}\vec{i} + \frac{\partial \varphi}{\partial y}\vec{j} + \frac{\partial \varphi}{\partial z}\vec{k}$$

(7)

what is possible write by equation (6) as

$$\vec{E} = -grad\ \varphi$$

(8)

where $-grad\ \varphi$ is vector in direction   .

Equation (8) enables determine electrical field intensity from known decomposition of electrical potential.

By substitution from equation (1) and (8) is created relation

$$d\varphi = grad\ \varphi\, d\vec{r}$$

(9)

If orbital element $\vec{dr}$ is perpendicular on vector $\vec{E}$ at once – and on *grad* φ, then potential will be constant and

$$d\varphi = 0$$

(10)

Grad φ determines direction of the highest growth of potential φ. Absolute value of *grad* φ is equal to increase of potential per angstrom in direction of vector $-\vec{E}$. Positive electrical charge can move in direction from lower to higher levels by the instrumentality of electrical forces analogous to fallen body in field of attraction.

By analogy we can described artificial potential field where in accordance to [1] is defined force field as negative gradient of potential that is associated to this field. Mathematical term of this formulation is:

$$\vec{F} = -\left\{\frac{\partial U}{\partial x}\vec{i} + \frac{\partial U}{\partial y}\vec{j} + \frac{\partial U}{\partial z}\vec{k}\right\} = -\nabla U$$

(11)

If is running any process in this field, what means that is stabilized, then is valid expression:

$$\nabla \vec{F} = 0$$

(12)

By using equation (11) is obtained expression on form:

$$\nabla \cdot \nabla U \equiv \nabla^2 U = 0$$

(13)

or

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = 0$$

(14)

The last equation (14) constitutes Laplace`s equation that is described behavior of arbitrary stabilized potential field.

B. *Harmonic potential field*

Harmonic potential field is possible define in accordance to [4] as a function on the set Ω as follows:

$$\nabla^2 U(x) = 0$$

(15)

with it that x belongs to the set Ω and ∇^2 is Laplace`s operator.

At method potential field is generated a virtual map of real space. Therefore is necessary to define bounds of space and bounds of obstacles. Border of the set Ω is defined as border of all obstacles. What does it mean? Every bordered space has his bounds and for robot these bounds of space are obstacles essentially. Therefore is necessary to add bounds of space in the set of bounds of all obstacles. Then is necessary define goal position also. From this knowledge are created conditions that potential field must satisfy. These conditions are in form:

1. $U|_\Omega = 0$
2. $U_{(goal)} = constant$

The first condition determines value for bounds of all obstacles of space, where belong bounds of space also. The second condition determines value for goal position.

At method harmonic potential field is needed to solve problem of partial differential equation of borderline values analogous to it is described in Dirichlet problem. Dirichlet problem is possible described as follows. All physical and technical problems that conduce to solving of Laplace`s equation are characterized that belong to borderline problems. Dirichlet problem belongs to these borderline problems also. In Dirichlet problem is imposed value of searched solution on boundary of set. At borderline problems is searched harmonic function. From principle of minimum and maximum results that Dirichlet problem is clearly, what means that Dirichlet problem has only one solution and this solution continuously depends on borderline conditions. We need to have only one minimum, global minimum at method harmonic potential field. Therefore, in this case, is used Dirichlet problem for solution of partial differential equations. These knowledge have a big applied meaning for solution of borderline problems.

By analogy form (11) is possible to determine relation for motion of robot in this potential field as follows:

$$\dot{x} = -\nabla U(x)$$

(16)

what we can described then as

$$\dot{x} = -\frac{\partial U}{\partial x} \qquad\qquad \dot{y} = -\frac{\partial U}{\partial y}$$

(17)

what is valid for two-dimensional system.

C. *Experiments*

By now we achieved some experimental results that are depicted on Figure 1, Figure 2 and Figure 3. We created the potential field that is simulated real space by using simulated software Student`s Quick Field. On the figures is depicted two-dimensional space without obstacles or with obstacles.

The goal position is only one minimum in this space. Therefore robot is always arrived to the goal position because his direction is every time controlled to this minimum from whatever position. Repulsive force has effect from both obstacles and bounds of space. Attractive force has effect to goal position.

On figures is depicted goal position with red lines, blue lines are limited bounds of space and border obstacle. By the instrumentality of attractive and repulsive force is created grid because every cell of space has a certain evaluation, goal position has the biggest evaluation. From figures we can see that every arrow represent one cell of space so space with this arrows is represented grid. Every arrow has direction to the goal position and from both obstacles and bounds of space.
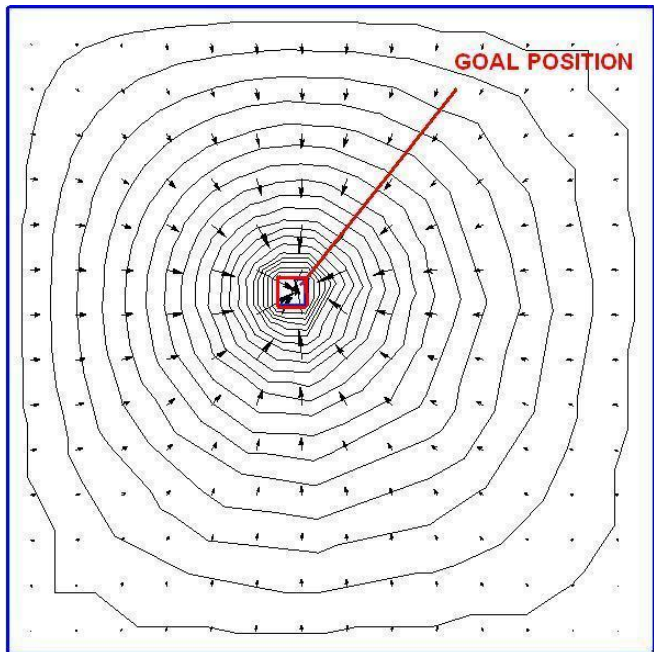
Figure 1: Created potential field without obstacles using Student`s Quick Field
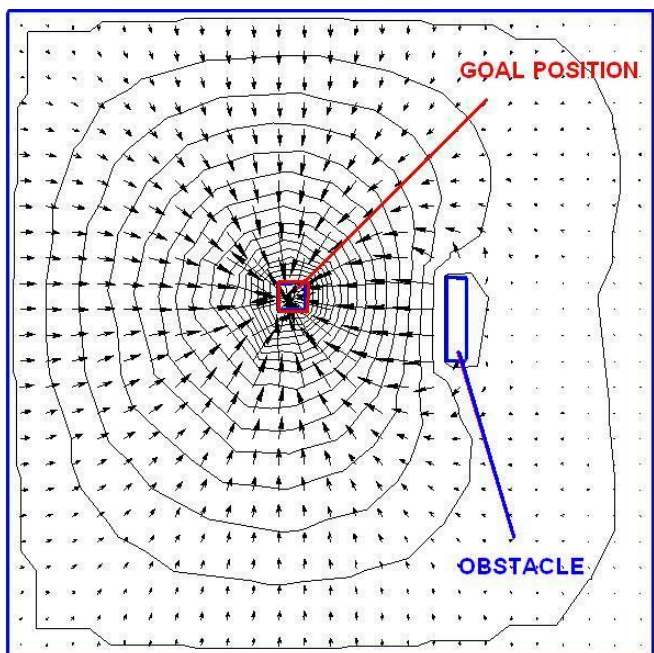
Figure 2: Created potential field with one obstacle using Student`s Quick Field
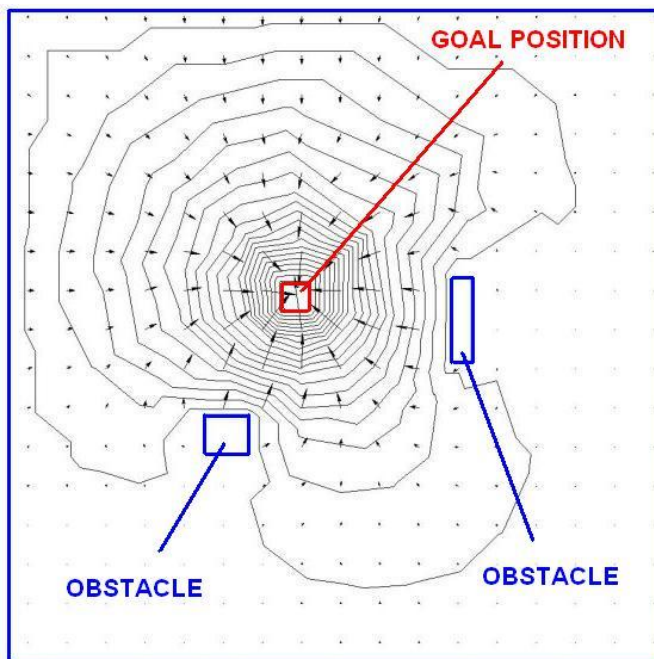
Figure 3: Created potential field with two obstacles using Student`s Quick Field

III. CONCLUSION

Method harmonic potential field is often used method for planning path of robots in space with obstacles. Mostly is used for static space what means that map of real space is known in advance. In this way were simulated our experiments where was known position of obstacles and position of goal. Harmonic potential fields are used for dynamic space also but it is lots more complicated.

REFERENCES

[1] Vaščák, Ján: Využitie potenciálových polí v navigácií mobilných robotov, Technická univerzita v Košiciach, 2008.
[2] Winkler, Zbyněk: Plánování na mřížce, 2003-12-03, dostupné na internete: <http://robotika.cz/guide/gridplan/cs>
[3] Siegwart, R. – Nourbakhsh, I.: Potential fields, October 16 2008.
[4] Masoud, A. Ahmad – Masoud, A. Samer: Motion Planning in the Presence of Directional and Obstacle Avoidance Using Nonlinear, Anisotropic, Harmonic Potential Fields, Mechanical Enginneering Department, Jordan University of Science & Technology, San Francisco, April 2000.
[5] Shahab, Mohammad: Harmonic Potential Field Motion Planning, King Fahd of Petroleum and Minerals, 7 February 2009.
[6] Hwang, K. Yong – Ahuja, Narendra: A potential Field Approach to Path Planning, IEEE Transactions on Robotics and Automation vol. 8 no. 1, February 1992.
[7] Ge, S. S. – Cui, Y. J.: New Potential Functions for Mobile Robot Path Planning, IEEE Transactions on Robotics and Automation vol. 16 no. 5, October 2000
[8] RNDr. Pavel Kubeš, CSc. – doc. RNDr. Zdeněk Kyncl, DrSc.: Fyzika I, České Vysoké Učení Technické v Praze, Fakulta elektrotechnická, Ediční středisko ČVUT, Praha, 1991
[9] Mayer, D. – Polák, J.: Metody řešení elektrických a magnetických polí, SNTL – Nakladatelství technické literatury, Alfa, Praha 1983

14

Frequency regulation process modeling in MODES

[1]Pavol HOCKO, [2]Marian HRINKO

[1]Department of Electric Power Engineering, FEI TU of Košice, Slovak Republic
[2]Department of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]pavol.hocko@tuke.sk, [2]marian.hrinko@tuke.sk

Abstract — **the paper describes the possibility of modeling supporting services in MODES. Synchronous model in which the two regulators are working is solved. Areas are connected by a single inter-tie line (Figure 1). Generator failure is simulated in the area. Load frequency control is monitored. Subsequently, the process of regulation is compared with different generator performance.**

Keywords — **frequency, MODES, load frequency control, regulation**
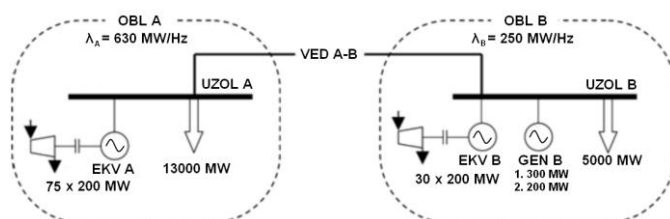
Fig. 1. Two modeled interconnected areas

I. INTRODUCTION

In complex interconnected power system are ongoing transition happens due to connecting and disconnecting loads and new sources to the system. For the stable operation of Power System (PS) is to be followed a power balance between production and consumption of electricity

The role of the Transmission System Operator (TSO) is to ensure observe this power balance and ensure operability of PS. Stability ensures TSO with system services. Ancillary services providing system services are divided in terms of time, to

- Primary active power control,
- Secondary active power control,
- Tertiary active power control,
- Remote voltage control,
- Start from blackout.

Primary active power control (or primary control) maintains the balance between production and consumption of electricity. Primary control operates in synchronous area with the speed regulators or active power control facilities providing ancillary services. [3, 55]

Secondary active power control balances changes to the planned tie-line exchanges with respect to frequency. Secondary active power control maintains the balance between production and consumption of electricity within each control area, respectively controlled prime mover. Programs of international exchanges of electricity are taken into account. Secondary active power control (or secondary control) works without disrupting operations of primary control.

Example process solves primary and secondary power control operation at different generator preferences. [2]

II. MODES

A MODES is a package of programs to analyze the dynamic behavior of complex power system. It is designed for power system transient calculation. A MODES works as DOS application on Windows 95 and higher. Work with the program makes it easy user interface MODMAN [4].

Synchronous area consisting of two regulatory areas was modeled in MODES. Areas are connected within inter-tie line on the voltage level of 400 kV (line VED A-B). The parameters of inter-tie line are in Table 2. It is a 50 km long line formed by bundled conductors. Circuit of interconnected system is shown in Figure 1.

Regulatory Area A is simplified to a single node to which supplies 75 generators of 200 MW. Regulatory Area B consists of 30 generators of 200 MW and B GEN source of 200 MW (or 300 MW for the second case). System parameters are in Table 1. We monitor the regulatory process after failure of the generator BLOK B [1]. Table 3 shows the parameters of the units participating in the system. Value N_{min} and N_{max} are the regulatory range of the turbines.

TABLE I AREAS PARAMETERS

Name	Frequency response char.	Load power	Generation power	Installed capacity	Number of sources
OBL A	630 MW/Hz	13000 MW	13000 MW	15000 MW	75
OBL B	250 MW/Hz	5000 MW	5000 MW	6000 MW	30

TABLE II INTER-TIE LINE PARAMETERS

Sending node	Receive node	R (Ω/km)	X (Ω/km)	B (μS/km)	Conductor type	Length (km)
NODE A	NODE B	0,023	0,279	3,98	3x450 AlFe	50

TABLE III GENERATOR PARAMETERS

Generator name	Node name	Number of sources	Source Sn (MVA)	Turbine Nmin (MW)	Turbine Nmax (MW)
EKV A	NODE A	75	-	100	200
EKV B	NODE B	0	-	100	200
GEN B	NODE B	-	231	220	400

In MODES is possible to specify a unit as a set of n - generators (units A EKV and EKV B) or as a separate source. If is source defined as individual, apparent power in MVA is entered. Apparent power should be converting through the generator power factor. In our case, the machine power factor is 0.866. Thus, apparent power generator is:

$$S_N = \frac{P_N}{\cos\varphi_N} = \frac{200}{0,866} \cong 231\,MVA \qquad (1)$$

Modeling activity of primary and secondary active power regulation will take place as follows. At t = 30 s is the case scenario set to disconnect the generator GEN B. Disconnect power supply causes imbalance in the area B. The frequency changes throughout the interconnected system. To frequency change response primary power control, this activates all sources in interconnected system. Secondary control active power continues independently of the primary control. Secondary control increases power resources in the system where there was a failure. We monitor compliance with the technical requirements for primary and secondary control power output regulation as defined in document [3]. Technical requirements defined by the TSO. Course of active power units, output of the turbine, the deviation from the nominal frequency of 50 Hz and the inter-tie line power flow within two cases is compared.

Two cases are modeled. The first case considers the failure unit named GEN B, with a rated active power of 200 MW. In the second case, for comparison, unit GEN B has 300 MW rated power. The individual graphs always compare selected variables for both cases. Thus, it is possible to compare the process of primary and secondary active power control at a different power source generation outages. We also monitor the amount of deviation of frequency and time of regulation activity. Simulation time was set to 900 seconds, which is 15 minutes. The sampling period is 1 second.

III. REGULATION PROCESS

A. Prime mover

In Figure 2 the performance of prime movers participating to primary active power control in both areas are compared. Values are shown in per unit. Performances are compared for both cases. First case is failure of 200 MW units (blue waveforms). Second case is failure of 300 MW units (red waveforms). We can monitor activity of the primary active power regulation on waveforms. Primary active power control should works in all controlled areas in interconnected system.

Both turbines improve their performance. Figure 2 shows that the primary active power control is working properly.
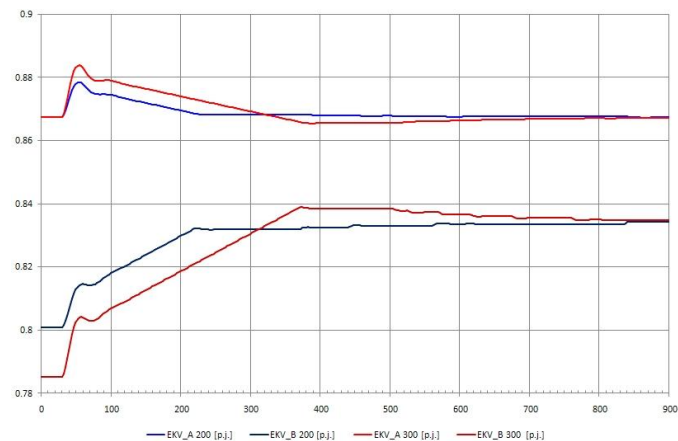
Fig. 2. Turbine performance waveforms for both cases in per unit

Secondary power control took over from the primary power control. The secondary active power regulation is involved only in areas where there has been a failure. In the work, of secondary power control the performance to block EKV A gradually decreases. Failure is the offset in the area where it occurred. In the Figure 2 is possible to compare the different waveforms activity in the regulation of generation outages with different sizes of power.

B. Generators performance

Figure 3 shows a plot with the course of active power sources in the interconnected system. Active performance supplied by generators depends on the power turbine. The Figure 3 shows a transition effect resulting from the disconnection of the source of active power. Waveforms shown in blue are values in case of failure of 200 MW units. Similarly, as the previous figure, waveforms in case of failure of 300 MW are red. In the first case (blue waveforms) we follow the smaller contributions of resources to compensate for the loss of power. Primary active power regulation works in both systems. Increase the output power generators to offset the power balance. The contribution of primary control is proportional to size failure. Secondary active power control works only in areas where there has been a failure. In our case it is the source named EKV B. active power is increased in power system where there has been a failure. Activity on the primary power control follows the secondary power control. Secondary power control must be activated no later than 30 seconds from the end of the fault and must work within 15 minutes. Figure 3 shows the correct action of the secondary active power control in both referred cases. Primary active power control works also correctly.

14
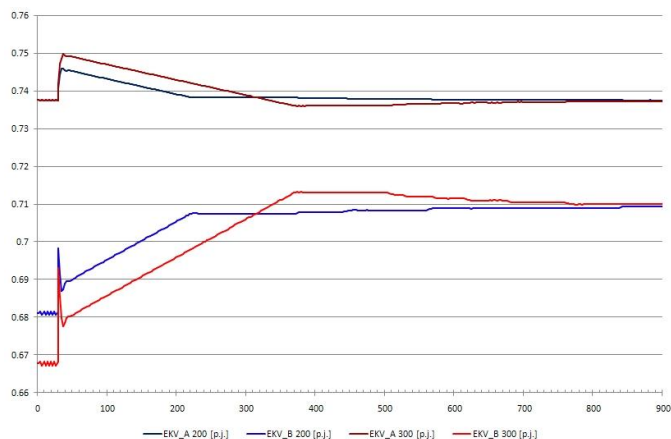
Fig. 3. Generators performance waveforms for both cases in per unit

C. Frequency deviation and power balance

Figure 4 shows a plot with system balance in the B power system and the frequency deviation from the nominal value of 50 Hz. Also, frequency deviation is shown in Figure 5 separately. Frequency deviation waveforms are shown in per cent. Again, both cases are compared. Failure on the source of 200 MW is again shown in blue. The second case is shown in red. Since the areas are interconnected within only one inter-tie line the balance value is the same as the power delivered by inter-tie line VED A-B. In the time from zero to 30 seconds, we can observe zero transmitted power. This means that in both areas before the disturbance was maintained power balance and inter-tie line was transferring no power.

At the time of failure will sharply increase the transmitted power. It also dramatically changes the frequency offset, the maximum value is referred to as dynamic deviation. Primary power control responds to the frequency deviation and increase sources performance. Size of power transmitted trough inter-tie line is equal to the contribution of primary sources participating in the of regulation power in system A. Resulting deviation of the frequency offset secondary power control. Action in the secondary power control, system B (where was a failure) reduces the transmitted power to zero.

The frequency deviation is also reduced. Comparing both cases, we can follow a different intensity of action of regulation. Frequency deviation is more significant after a failure generator with more power. In 300 MW unit fail is secondary active power regulation more significant and action time is longer than in case of 200 MW power supply failure.
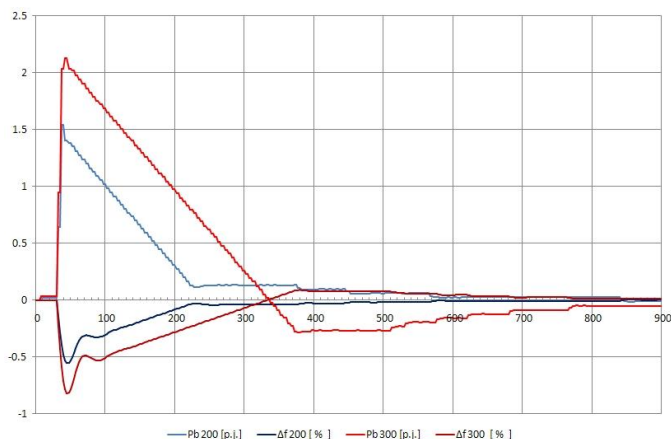
Fig. 4. Power balance waveform for both cases shown in per unit

The controller does not work exactly. The controller increases the power of the source of value needed to achieve power balance. But the compensation frequency deviation occurs in the required time limit 15 minutes (900 seconds)
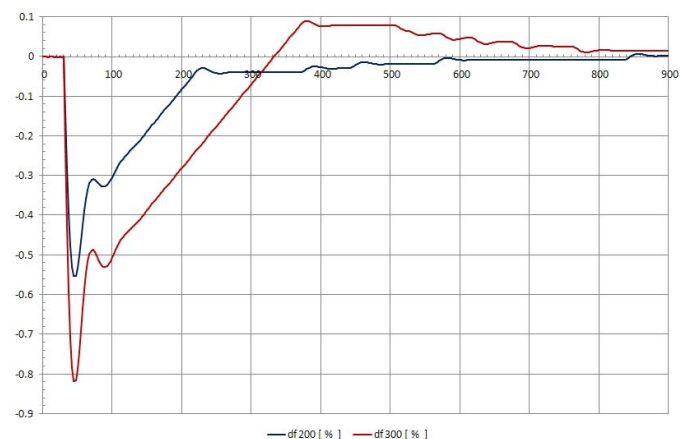
Fig. 5. Frequency deviation waveforms for both cases shown in per cent

IV. CONCLUSION

The process of primary and secondary power control depends on the size of the power outage. Primary power control responds only to change the frequency deviation. The frequency change is directly dependent on changes in power generators or sources. Since then, it depends on the action the secondary power control. Time operation performance of secondary power control regulation is dependent on the size of the frequency deviation which remains after the action of the primary power control regulation. Controller parameters were the same in both cases. Performance indicator system is unchanged, and therefore compensation of the frequency from secondary power control has the same course with an equivalent delay. Frequency deviation values are significantly larger than the deviation which would lead to failure in the real interconnected system. It is caused by a model in which only two power systems are interconnected. The aim of simulation was to show the dependence of frequency variations when changing the active power. The frequency and power control worked basically correct.

REFERENCES

[1] HOCKO, Pavol, Diplomová práca, Modelovanie podporných služieb v programe MODES

[2] KOLCUN, Michal – GRIGER, Vojtech – BEŇA, Ľubomír – RUSNÁK, Jozef: Prevádzka elektrizačnej sústavy. Košice: TU-FEI, 2007, ISBN 978-80-8073-837-2

[3] SEPS a.s. : Technické požiadavky na pripojenie prístup a prevádzkovanie prenosovej sústavy. Dokument B. Technické podmienky pre pripojenie a prístup k prenosovej sústave.

[4] MASLO, Karel: Dokumenty k programu MODES, Dlouhodobá dynamika

Hybrid FSO/RF Communication System

[1]Pavol MIŠENČÍK

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]pavol.misencik@tuke.sk

Abstract— **Free-space optical (FSO) links offer gigabit per second data rates and low system complexity, but suffer from atmospheric loss due to fog and scintillation. Radio-frequency (RF) links have lower data rates, but are relatively insensitive to weather conditions. Hybrid FSO/RF links combines the advantages of both links. This article describes the challenges of FSO/RF hybrid system, the impact of environment on this system and the effectiveness of these hybrid systems.**

Keywords— *Free-space optical (FSO), Hybrid FSO system, Optical wireless, Hybrid network.*

I. INTRODUCTION

For last 10-15 years Free Space Optics (FSO) has become from an experimental laser technology to a viable, high-bandwidth wireless alternative to fiber optic cabling [2]. The primary advantages of FSO over fiber cable communications systems are its easy installation, rapid deployment time and significant cost savings. The disadvantage of FSO systems over fiber systems is variable laser power attenuation through the atmosphere which it is very dependent on weather. The atmosphere absorption significantly limits the distance at which FSO system should be deployed with projected factor of reliability. The FSO system integrators use historical weather data that are usually collected at airports to predict the link availability as a function of distance. These weather data provide a good indication of the reasonable link distances for FSO systems in a particular geographical area. FSO link distances can vary greatly from desert areas to heavy-fog big cities, ocean foggy areas. Another factor in determining FSO link distance limitations is the link availability expectation of the application.

For enterprise applications, link availability requirements are generally 99,9% or greater [2,3]. This allows for relatively longer FSO link ranges, based on the weather prediction. The enterprise market is where the majority of FSO systems have been deployed. For carrier-class the FSO links availability is generally considered to be 99.999% (5 nines) [2].

Thus for telecommunication applications, they have to meet much higher availability requirements than for enterprise ones. The carriers and ISPs are another potential large user of high-bandwidth FSO systems, either for wireless backhaul or last-mile access applications. One such of solution is back-up FSO link with RF link which maintains a connection. This paper describes the FSO/RF system that can maintain the required reliability [2].

II. THE HYBRID FSO/RF SYSTEM

A. Description of system

The experimental system is designed to make use of the best features of two transport mediums, laser light and radio waves to form a single, high available, seamless wireless communication link between the two campuses. As shown in Fig. 1, the system comprises three subsystems namely, the laser link, a switch and the RF link at each end.

Under normal operating conditions data is transmitted over the laser link where the laser link provides a 155 Mbps full duplex connection. The wavelength of the laser used is 1550 nm. Other available wavelengths for FSO applications are 780 nm and 850 nm near infrared spectrum. However, 1550 nm band, also the choice for the fiber optic telecommunication applications, is better suited for optical wireless. The main benefit of the 1550 nm band is the ability to transmit more power. The power density at 1550 nm band nearly 50 times
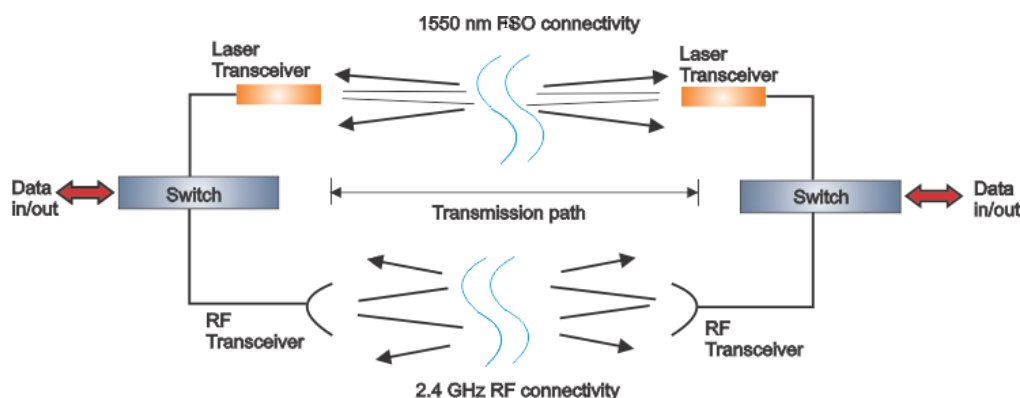
Fig. 1. Basic hybrid FSO/RF system architecture

that at 780 nm and is still safe for the human eye. Since, more power means better ability to overcome atmospheric attenuation, the 1550 nm band has been chosen [1].

Despite the more power, in heavy attenuation the laser signal might still get degraded. The signal level at the receiver gets checked every 5 seconds and when the received signal level falls below a certain threshold, the laser stops transmitting data to the switch. As soon as the degradation at the signal level is detected, the switch automatically switches over to the RF link, which is waiting as a hot stand-by.

Thus, despite the reduction in the bandwidth, the link remains available. The RF link operates at 2.4 GHz linking the two terminals at 11 Mbps. After the switchover has happened the RF link stays as the primary link until the received signal level goes over a higher threshold, indicating a solid connection on the laser link can be re-established. When the increased signal level is detected the switch switches back to the laser link.

B. Design considerations

The system has been designed to communicate through the FSO link unless the performance of the FSO is degraded due to adverse weather conditions. When the degradation occurs and the quality of the communication reduces the system switches over to the RF link via a switch. A threshold level on the received laser power must be determined for a seamless switchover to the RF link. In threshold level calculations, the gains and losses of optical components have been neglected and a 3 dB margin has been added on top of the calculated level instead in order to account for then neglected optical component losses. Even when the system switches to the radio, the received power level is still monitored for a quick recovery of the laser link so that a much higher bandwidth could be utilized. However, the system does not switch back over to the laser link even when the FSO-to-RF switchover threshold power level is achieved again. In fact, the power level to go back on the FSO is higher than the FSO-to-RF threshold. The reason for having two different switchover thresholds (a higher one for switching back to the laser from the radio than the one to switch from the laser to the radio) is to reduce the switchover frequency and prevent the system switching back and forth constantly [1].

C. Path reconfiguration

A switching process occurs as a result of loss of an active link caused by one or more of reasons, e.g. loss of line-of-sight, atmospheric conditions, or device failure. During such an event, the next path is chosen, and, data are re-routed. This section details describe the path reconfiguration Pro algorithm. The transmitter broadcasts data through FSO using the direct path with the highest priority. In case of a link loss during data transfer or when the received signal level falls below a certain threshold, the laser stops transmitting data and switch on RF path. The transmitter continues data transfer through RF path. This is how an alternate path is chosen. Fig. 2 shows detailed of description algorithm. It should be noted that the network attempts to transmit data using FSO as long as possible. However, this is not possible if none of the FSO paths formulated in the topology discovery process are available. In the absence of a FSO link, data transmission occurs through the direct RF path between the transmitter and receiver. Clearly, the consequence of switching to the RF

network occurs only after all FSO paths are attempted and unavailable [3].
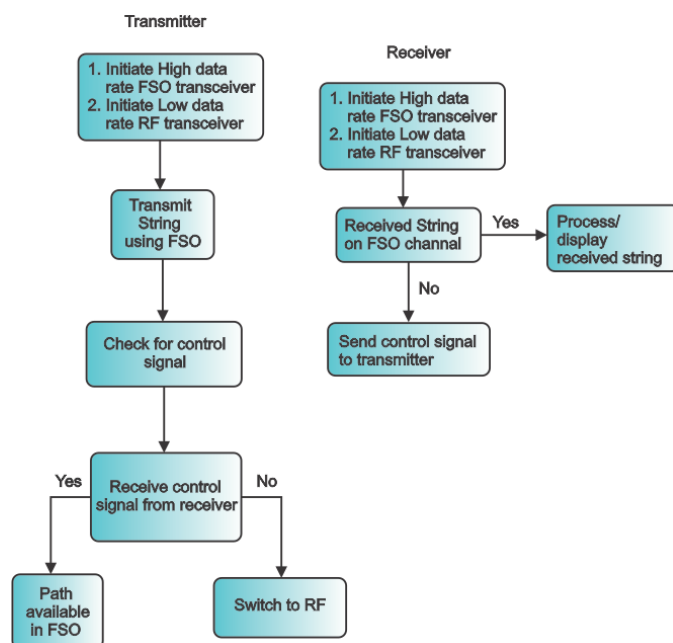
Fig. 2. Switch algorithm

III. CONCLUSION

FSO is the most promising candidate among different communication technologies when a high data rate link is required for a particular terrestrial and ground-space application thanks to its notable prime advantages. Despite its major advantages, this access technology suffers from severe availability and reliability challenges mainly due to different weather effects like fog, rain, snow and clouds in the earth atmosphere. Hybrid network, a combination of FSO main link and radio frequency back up link, is a viable solution to overcome the atmospheric attenuations. This paper presented an experimental hybrid FSO/RF communication system.

REFERENCES

[1] AKBULUT, A., EFE, M., CEYLAN, M. A., ARI, F., TELATAR, Z., ILK, H. G., TUGAC, S.: An experimental hybrid FSO/RF communication system. In: Software, Telecommunications and Computer Networks, 2008, pp. 356 - 361

[2] The Hybrid FSO (HFR) concept: Carrier and enterprise class FSO link back-up by ELVA 100Mbps wireless LAN bridge with 100Base-T interface.
http://www.elva-1.com/products/telecom/FSO_mmW_backup.

[3] GURUMANI, S., MORADI, H., REFAI, H. H., LOPRESTI, P. G., ATIQUZZAMAN, M.: Dynamic Path Reconfiguration among Hybrid FSO/RF Nodes. In GLOBECOM, 2008, pp. 1-5

Induction Motor Vibration Power Analysis

[1]Oksana BRATASH, [2]Andrey KALYNOV,[3]Anna HODULÍKOVÁ,

[1]Automated Control Systems and Electric Drives department, Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine
[2]Automated Control Systems and Electric Drives department, Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine
[3,4]Dept. of Theoretical Electrical Engineering and Electrical Measurement, Technical University of Košice, Slovakia

[1]bratashka2009@mail.ru , [2]scenter@kdu.edu.ua, [3]anna.hodulikova@tuke.sk,

Abstract — **Based on the analysis of existing methods for determining the vibration power of the electrical machine, the method of analytical calculation of the vibration power has been suggested. It is based on the algorithm of the discrete convolution of two series that are represented as vectors of the vibration velocity and acceleration. It is shown that the vibration power of the induction motor (IM) may consume a significant portion of active power, it should be considered during technical diagnostics.**

Keywords – **vibration-power, vibration, vibration monitoring.**

I. INTRODUCTION

Energy conversion in electrical machines and technological mechanisms is invariably associated with variable forces and moments causing the vibration. When there is some damage in the electric machine, the vibration level rises. Losses caused by the increase in the vibration power make an effect on the total loss of consumed power, as they are a part of the overall energy balance of the engine.

The spectral and phase characteristics of vibration displacement, vibration velocity and vibration acceleration are separately used for analysis of vibration. They do not show the distribution of energy flows in the tested object. The investigation of the vibration power of the electromechanical system makes it possible to analyze the distribution of these flows in the time and frequency domain [1].

The vibration power should be considered as a main characteristic in the evaluation of the machine vibration activity in the specified operating conditions.

The results of measurements of outcomes vibration-power can be used for:

- identification and comparison of vibration activity of machines of various designs, the mass and dimensions in specific operating conditions;
- development of technical requirements for permissible levels of machine vibration;
- development of solutions for machine vibration monitoring and its procedure;
- solving the problems of the diagnostics of machines vibration acoustic and vibration acoustic protection;
- development of measures for reducing the vibration activity of machines and increasing the effectiveness of vibration acoustic protection of machines [2], [3].

II. ANALYSIS OF EXISTING METHODS CALCULATION OF VIBRATION POWER

Existing approaches to measuring and calculating the vibration power are described in the publications [2]-[4]. Instant vibration-power is obtained on the basis of vibration velocity and vibration acceleration signals, measured in the desired direction. In general case, the instantaneous vibration-power calculated as follows:

$$p_v(t) = v_v(t)f(t), \text{ W} \qquad (1)$$

$v_v(t)$ - instantaneous value of the vibration velocity along one of the three axes, mm/s; $f(t)$ - instantaneous value of force directed at one of the axes, N.

In [2], [3] a method of indirect measurements of vibration power is described. The power is determined according to the analysis of the joint spectra of vibrations on the input and output of vibration isolator, and on the basis of information about the mechanical resistances of machine vibration isolators.

$$p^n{}_v(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{t1}^{t2} f^n(t)v_v(t)dt, \qquad (2)$$

$\Delta t = t2\text{-}t1$ - the time interval over which the averaging is performed, n - number of the vibration isolator.

In [4] the calculation of vibration power is based on the following expression:

$$p_v(t) = v_v(t)a_v(t)m_r, \qquad (3)$$

$a_v(t)$ - the instantaneous vibration acceleration along one of the three axes, m/s²; m_r ,- reduced mass, kg.

The advantage of this method is in no need for force sensors, only the vibration sensor is used. Only one parameter is directly measured: vibration velocity or acceleration. The second parameter is calculated by the indirect method based on the integration or differentiation of the original signal in time:

$$v_v(t) = \int a_v(t)dt; \qquad (4)$$

$$a_v(t) = \frac{dv_v(t)}{dt}. \qquad (5)$$

To convert the acceleration into the velocity, the integration should be applied. It requires more complex equipment that results in additional errors. Moreover, the measurement accuracy of low-frequency vibration may decrease due to the fact that these measurements are carried out near the lower scope of resolving ability of devices that inevitably leads to distortion of the results [2].

Based on the analysis we suggest developing a new method for measuring and calculating the vibration power, where lacks of existing methods are taken into account.

III. MATHEMATICAL APPARATUS OF CALCULATING VIBRATION-POWER

A. Calculation of harmonic components of vibration velocity

As mentioned above, the most rational way to determine the vibration power is a technique in which only one of vibration parameters are measured directly. Since the most frequently used instruments for measuring the vibration is the accelerometer, then the vibration acceleration can be used as a directly measured signal.

The errors of instantaneous vibration velocity calculation can be avoided if instead of the instantaneous value we use the harmonic composition of the vibration acceleration. In this connection the operation of integration is replaced by the calculation of harmonic components of vibration velocity $V_{V\,k}$ across the harmonic components of vibration acceleration $A_{V\,k}$:

$$V_{V\,k} = \frac{A_{V\,k}}{k/T \cdot 2\pi}, \text{ mm/s,} \quad (6)$$

T - time interval at which the signal is analyzed, s; k - number of harmonics in the expansion of acceleration signal in a Fourier series

B. Determination of complete instantaneous vibration-power

From Newton's second law, the expression for calculating the total vibration power is as follows:

$$P_{V\,k} = \left[m_r A_{V\,k} + m_r g \right] V_{V\,k}, \quad (7)$$

m_r – Reduced mass, kg; g - Free fall acceleration, m/s^2.

In addition to the dynamic component of the vibration power transmitted to the supports, the method of determining which is described in [2]-[4], there is a static component $P_{V\,st\,k} = m_r g V_{V\,k}$, caused by the machine weight effect on the supports. It should be taken into account.

C. Algorithm for calculating the instantaneous components of vibration-power

To calculate the instantaneous vibration power, based on orthogonal harmonic components of vibration velocity and vibration acceleration we use the algorithm [5], [6], which is based on the theorem of convolution of the discrete series. The algorithm is implemented in the MathCad

$$P_{\hat{a}} := \begin{vmatrix} \begin{array}{l} \textit{for } m \in 0, 1..M \\ \left| \begin{array}{l} P_m \leftarrow 0 \\ \textit{for } k \in 0..N \\ \left| \begin{array}{l} P_m \leftarrow P_m + ReI_k \cdot Re \cdot U_{m-k} - ImI_k \cdot Im \cdot U_{m-k} \quad \textit{if } m\text{-}k \geq 0 \\ P_m \leftarrow P_m + ReI_{|m-k|} \cdot ReU_k + ReI_k \cdot ReU_{k-m} + ImI_k \cdot ImU_{k-m} + ImI_{|m-k|} \cdot ImU_k \quad \textit{otherwise} \end{array} \right. \\ P_m \leftarrow \frac{P_m}{2} \textit{ if } m\text{=}0 \\ P_m \leftarrow \frac{P_m}{2} \textit{ if } m\text{>}0 \end{array} \right. \\ P \end{array} \end{vmatrix} ,$$

$$P_b := \begin{vmatrix} \begin{array}{l} \textit{for } m \in 0, 1..M \\ \left| \begin{array}{l} P_m \leftarrow 0 \\ \textit{for } k \in 0..N \\ \left| \begin{array}{l} P_m \leftarrow P_m + ImV_k \cdot ReA_{m-k} + ReV_k \cdot ImA_{m-k} \quad \textit{if } m\text{-}k \geq 0 \\ P_m \leftarrow P_m + ImV_k \cdot ReA_{k-m} + -ImV_{|m-k|} \times ReA_k + ReV_{|m-k|} \cdot ImA_k + -ReV_k \cdot ImA_{k-m} \quad \textit{otherwise} \end{array} \right. \\ P_m \leftarrow \frac{P_m}{2} \textit{ if } m\text{=}0 \\ P_m \leftarrow \frac{P_m}{2} \textit{ if } m\text{>}0 \end{array} \right. \\ P \end{array} \end{vmatrix} ,$$

Pa - cosine component of the instantaneous vibration-power; Pb - sine component of the instantaneous vibration-power; ReAk, ReVk - cosine components of the vibration acceleration and velocity; ImAk, ImVk - sine components of the vibration acceleration and velocity;
N, M - quantity of harmonics of the vibration acceleration and velocity, respectively; m, n - number of harmonic of the vibration acceleration and velocity.

package using the symbolic computation subsystem.

IV. EXPERIMENTAL RESEARCH

A. *Experimental conditions*

For the numerical analysis of electric engine vibration power by the proposed method the experimental studies of vibration signals were carried for MF-11-16 -type engine with the parameters of Pn = 1.4 kW, In = 9.3 A, n = 886 rpm, m = 70 kg [7].

The table 1 shows the most typical experiences at work IM from the symmetric (K(unsymmetry) = 1,1%) and asymmetric (K(unsymmetry) = 20,8%) power supplies. The analysis is

appears when levels of vibration acceleration, vibration velocity, vibration power and consumption power are increased, especially it is typical for the axial direction (axis Z).

During the study of the vibration power spectrum formation, the calculations were carried out by the full spectra and selected, the most important, components of the spectra of vibration velocity and vibration acceleration. When comparing the spectra of original and recalculated vibration power, it has been found that in the first case, RMS = 50,63W, and in the second - RMS = 34.325 W.

These results show that the vibration power RMS is significantly affected with the intermediate, not major,

TABLE I
RESULTS OF EXPERIMENTS

Symmetrical mode												
RMS	V_v, mm/s				a_v, m/s^2				P_v, W				
	12.5-87.5 Гц	87.5-400 Гц	400-1000 Гц	12.5-1000 Гц	12.5-87.5 Гц	87.5-400 Гц	400-1000 Гц	12.5-1000 Гц	12.5-87.5 Гц	87.5-400 Гц	400-1000 Гц	1000-3000 Гц	Реф
X	32.81	2.738	3.174	33.08	5.131	2.561	13.61	14.77	9.539	4.505	23.55	5.747	26.44
Y	46.73	1.993	1.666	46.79	4.995	2.685	7.121	9.731	9.569	6.7	17.2	15.58	26.03
Z	53.47	2.672	1.402	53.55	6.217	4.086	5.757	9.407	17.114	11.14	15.72	7.529	26.89

P, W						
$P_{31.25}$	P_{50}	P_{100}	P_{200}	P_{300}	P_0	$P_{эф}$
18.64	27.95	143.55	13.57	150.94	935.56	244.47

Asymmetrical mode												
RMS	V_v, mm/s				a_v, m/s^2				P_v, W				
	12.5-87.5 Гц	87.5-400 Гц	400-1000 Гц	12.5-1000 Гц	12.5-87.5 Гц	87.5-400 Гц	400-1000 Гц	12.5-1000 Гц	12.5-87.5 Гц	87.5-400 Гц	400-1000 Гц	1000-3000 Гц	Реф
X	13.28	6.765	5.602	15.89	2.814	9.489	23.44	25.36	3.126	9.428	19.23	8.52	23.21
Y	14.27	6.663	5.972	16.77	2.487	13.43	22.86	26.33	4.012	14.00	19.84	16.40	29.45
Z	16.36	28.49	5.179	33.11	3.459	26.89	18.07	31.66	16.283	42.70	39.00	13.30	61.43

P. W						
$P_{15.625}$	P_{50}	P_{100}	P_{200}	P_{300}	P_0	$P_{эф}$
26.105	48.48	$2.04 \cdot 10^3$	389.59	195.56	$3.21 \cdot 10^3$	$2.14 \cdot 10^3$

performed by three spatial axes (X, Y, Z). The assessment of vibration power is conducted by the parameter of RMS for four frequency ranges: 12.5-87.5 Hz - low-frequency vibration, 87.5-400 Hz and 400-1000 Hz - vibration in central frequency band 1000-3000 Hz - high frequency vibration, and for the entire frequency range (Table 1). Such a separation is necessary because the reasons causing the increase in vibration on these frequency ranges are of different nature.

B. *Results of experiments*

The study of electromagnetic and mechanical vibrations for these modes shows that the asymmetry of power supply

harmonic components of vibration acceleration and vibration velocity. The analysis of experimental data shows that for IM symmetric mode the total vibration power RMS by three axis is 79,37 W, ie 8.2% of the average three-phase power consumption. For the asymmetric mode the total effective vibration-power 114,1 W, ie 3,55 % of the average power consumption. With respect to the nominal electric power of IM, the total effective vibration power is 5,67% and 8,16% respectively for symmetric and asymmetric modes.
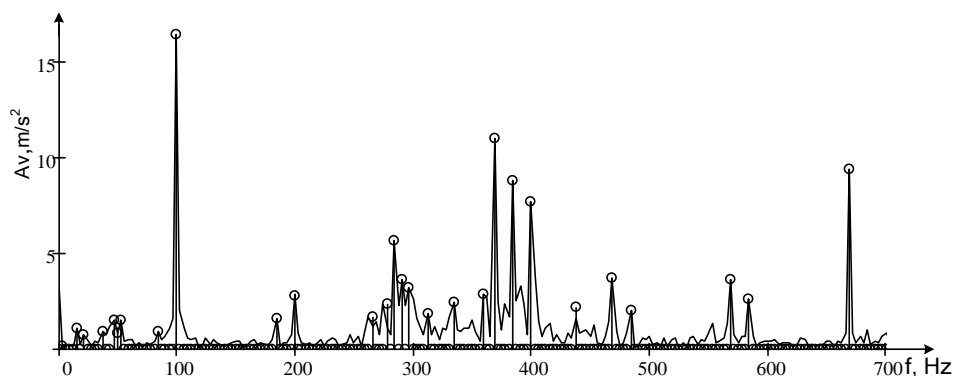
Fig. 1 Spectrum of vibration acceleration along the Z axis with phase unbalance

(——) - The original signal, (○) - basic harmonic components
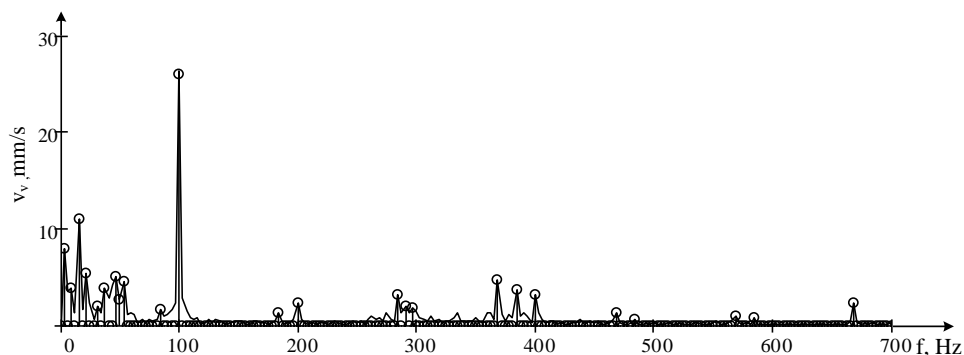
Fig. 2 Spectrum of vibration velocity along the Z axis with phase unbalance

(——) - The original signal, (○) - basic harmonic components
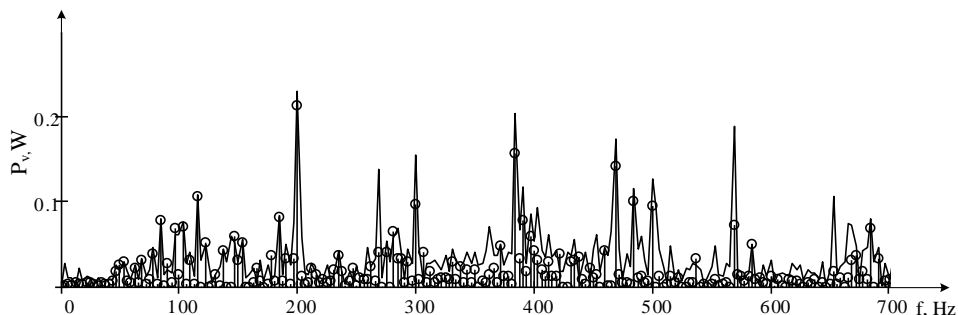
Fig. 3 Spectrum of vibration-power Z-axis with phase unbalance

(——) - The original signal, (○) - basic harmonic components

V. CONCLUSION

The analysis of the data shows that the vibration power makes up a significant part of the IM power consumption that should be taken into account in the overall energy balance when performing engines diagnostics. The complex analysis of the IM power consumption, the power of heat losses in its components, and the vibration power make it possible to evaluate the distribution of energy flows of different physical nature. The developed method allows us to estimate the energy component of the vibration power for diagnostics and assessment of life time of IM mechanical components.

ACKNOWLEDGEMENT

REFERENCES

[1] Вибрации в технике. Справочник. В 6-ти т. / Ред. совет: В. Н. Челомей (пред.). – М.: Машиностроение, 1981. - Т. 5. Измерения и испытания. – Под ред. М. Д. Генкина. - 496 с.

[2] Вибрация энергетических машин. Справочное пособие. [Под ред. Григорьева Н. В.] –Л.: Машиностроение, 1974. – 464 с.

[3] Вибромощность, излучаемая машинами в присоединенные опорные конструкции. Часть 2. Методика выполнения косвенных измерений: ГОСТ Р 8.653-2009. – [введен в действие 01.01.2010].– «СТАНДАРТИНФОРМ», Москва, 2009. – 27 с.

[4] Устройство для измерения мгновенной мощности механических колебаний: пат. №2189566 C2, Рос.: МПК G01H1/00 / Мартынов В. И., Федяев В. Л., Иванов Д. Ю.; /ЮУрГУ/ заявка №2000119330/28/, опубл. 20.09.2002.

[5] Калінов А. П. Автоматизований метод розрахунку електричних кіл за складовими миттєвої потужності /[А. П. Калінов, М. С. Малякова] // Електромеханічні і енергозберігаючі системи.– Кременчук: КДУ, 2009. – Вип. 1(5). – С. 34-38.

[6] Сидоренко В.М. Автоматизація розрахунку складових миттєвої потужності електричних сигналів / [Сидоренко В. М., Родькін Д. Й, Чорний О. П., Мамчур Д. Г.] // Вісник КДПУ: Наукові праці КДПУ. – Кременчук: КДПУ, 2004. – Вип. 3 (26). – С. 91–96.

Investigation of discharge activity in different types of grounding of high voltage coil

[1]Milan KVAKOVSKÝ, [1]Lýdia DEDINSKÁ, [1]Vieroslava SKLENÁROVÁ

[1]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]milan.kvakovsky@tuke.sk, [2]lydia.dedinska@tuke.sk, [3]vieroslava.sklenarova@tuke.sk

Abstract—**Despite improving the quality of insulation, a sizeable proportion of failures of electrical machines coupled with the right insulation failure. To prevent such failures can be consistent adherence to operating parameters and improved maintenance. However, one of the phenomena, which have a dominant influence on the quality of insulation, can only affect the least. It is the aging of insulation, which is linked to degradation factors affecting the very structure of the insulator. By changing its properties, the stability of certain tolerances for the operation of electric machine is necessary.**

Keywords—**insulation, partial discharges, slot, electrical machine.**

I. INTRODUCTION

In terms of diagnosis is the most watched element of high voltage rotating machine stator winding insulation, which is exposed during operation of a combination of various stresses. This is especially the electrical and thermal stress, but in no small measure to isolate the active mechanical stress that is caused by vibration of the machine. All types of stress adversely affect the insulation system of stator windings [1].

In a high voltage electrical insulation of rotating machinery can be a small cavity filled with gas. In the manufacture of insulation systems is not always possible to avoid inhomogeneities (voids) in the dielectric. These cavities can arise, for example, in the production, degradation, isolation or as a consequence of large local electric stress. If we increase the voltage applied to a measured object, appear at a certain voltage in the cavities discharges having the characteristics of avalanche or small spark discharges. In areas with reduced electrical strength due to high intensity electric field in the insulation arise partial discharges [2].

Partial discharges are the result of local concentration of electric stress in the insulation or on the surface of the insulation and create a current (voltage) pulses with duration of less than 1μs. These discharges represent a breakdown in gas-filled cavities. In other words, partial discharges bridge the distance between two electrodes only partially and the rest of isolation can withstand wide operating voltage [3].

Partial discharges in insulation systems of stator insulation occur most frequently in these locations:
• in cavities in the copper coil and a major part of the insulation or cracks in the main insulation, which may be caused by delamination of the insulation system, called internal discharges,

• between the protective conductive layer (which covers the surface coil placed in the slot) and the wall of the slot, called surface discharge,
• at the output of coil from the slot due to inadequate management of the electric field semi conducting tape, called surface discharges at the outlet of the coil from the slot,
• on the surface of the coil, which is stored in the slot, when the electric field along the surface exceeds a critical value, called surface discharges in the slot [4].

II. TEST SAMPLES AND MEASUREMENT METHODS

To measure discharges activity were used high voltage coil (of induction motor with a rated output of 2.1 MW and the rated phase voltage of 3.6 kV) without semi conductive protection. The ends of the coils were connected with homogenization electrode.

Three measurements were made each with a different type of grounding:

a) The whole slot part of coil was wrapped with a conductive mesh (Fig.1.a)

b) The whole slot part of coil was wrapped with aluminum foil (Fig.1.b)

c) The slot part of coil was wrapped in four specific locations with aluminum foil with width of 5cm, and individual points were connected with electrically conductive mesh and routed to ground potential (Fig. 1.c).

The measured object was hanging on isolation lane in a Faraday cage.

For partial discharge measurement was used direct galvanic method. The measuring impedance is connected in parallel to the measured object. The advantage of this involvement lies in the fact that the breakdown of the measured object do not damage the meters.

III. READINGS AND DISCUSSION

To compare the measured results of all the measurements (from the voltage of the initial discharges to the associated nominal voltage of the coil for all three types of grounding) we have identified three voltage levels, for which we performed the evaluation (Table I, Table II, Table III.)
- Initial voltage of partial discharges,
- Nominal phase voltage of 3.6 kV,
- Rated associated voltage 6kV.
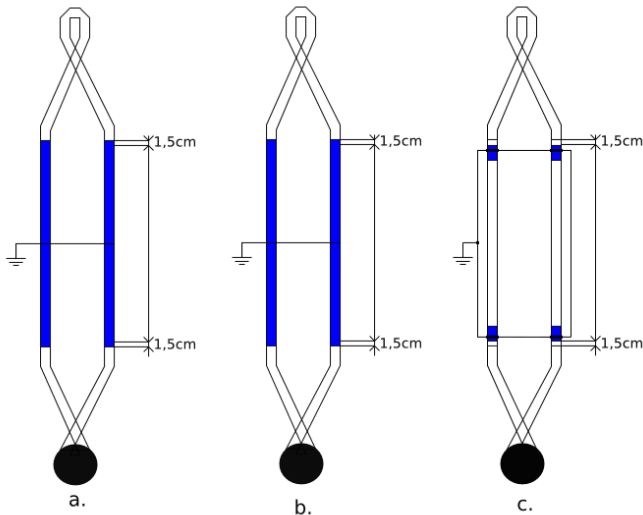
Fig. 1. Measurement of partial discharges in a variety of grounding, a.) ground through a conductive mesh, b) grounding by aluminum foil, c.) grounding by four points.

A. *Grounding of the slot part of the coil with conductive mesh*

Fig. 2 to 4 show the phase distribution of maximum apparent charge for the voltage levels of 3.4 kV, 3.6 kV and 6 kV. The voltage of the initial discharges in this type of grounding (conductive mesh) was 3.4 kV. The maximum value of apparent charge at this voltage reached 800 pC (Fig. 4) with a frequency of discharges of 0.02. Probably these discharges were the initial partial discharge, resulting in space between the mesh and insulation of the coil.

At a voltage of 3.6 kV discharge activity rose to 1800 pC (Fig. 5) with a frequency of discharges of 0.12. These are the surface discharges, resulting in the output of coil of stator slot, where is the largest electric field intensity.

The maximum value of apparent charge at a voltage of 6kV reached q_{max} = 20000pC (Fig. 6) with a frequency of discharge of 1.5. These are probably the surface discharge.

TABLE I
COIL GROUNDED THROUGH A CONDUCTIVE MESH

U [kV]	q_{max} [pC]	Nstr [-]	φ+ [°]	φ- [°]
3,4	800	0,02	40-70	230-260
3,6	2000	0,12	30-70	220-300
6	20000	1,5	20-80	210-310
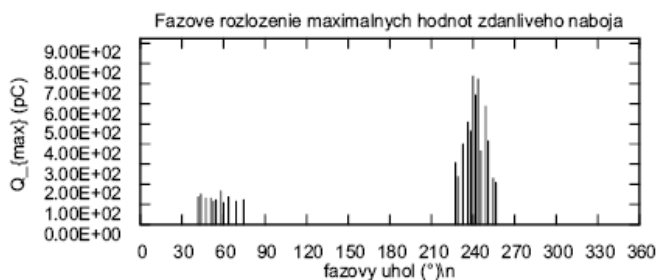
Fig. 2. Coil grounded through a conductive mesh at 3.4 kV
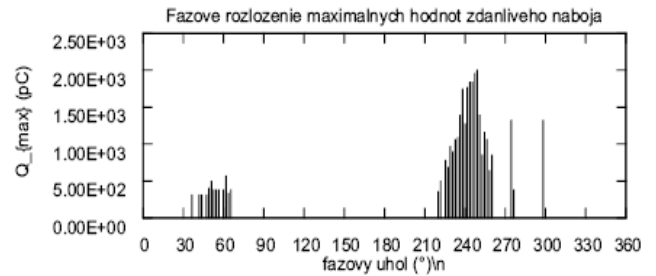
Fig. 3. Coil grounded through a conductive mesh at 3.6 kV
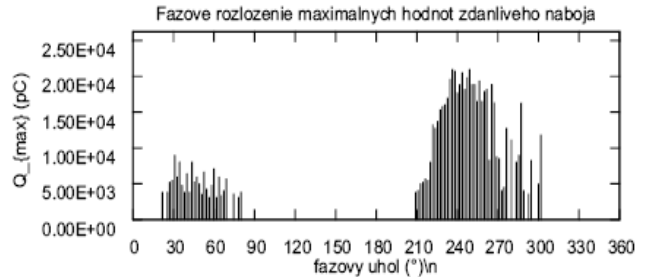
Fig. 4. Coil grounded through a conductive mesh at 6 kV

B. *Grounding of the slot part of the coil with aluminum foil*

Fig. 5 to 7 show the phase distribution of maximum apparent charge for the voltage levels 3kV, 3.6 kV and 6 kV. The voltage of the initial discharges in this type of grounding (aluminum foil) was 3 kV with the same frequency of discharge activity as before 0.02. The maximum value of apparent charge was 150 pC (Fig. 7). Discharges activity at this voltage level was unstable and gradually the discharges were decreasing.

At a voltage of 3.6 kV maximum apparent charge was 1200 pC (Fig. 8) with a frequency of discharges of 0.2. Compared with the previous case, the value of apparent charge is lower, the frequency of discharges increased. Discharges activity at this voltage level is represented by surface discharges.

At a voltage of 6kV discharges activity rose to q_{max} = 20000pC (Fig. 9). Even in this case it is only the surface discharge.

TABLE II
COIL GROUNDED THROUGH A ALUMINUM FOIL

U [kV]	q_{max} [pC]	Nstr [-]	φ+ [°]	φ- [°]
3,4	150	0,02	30-90	220-280
3,6	1200	0,2	30-90	210-300
6	20000	1,6	40-80	210-300
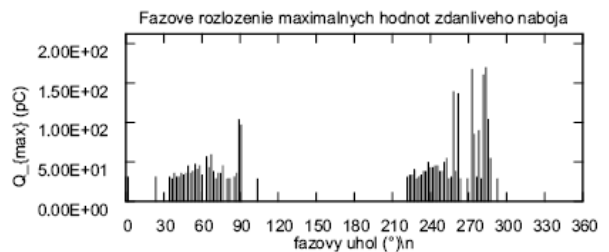
Fig. 5. Coil grounded through aluminum foil at 3.4 kV

14
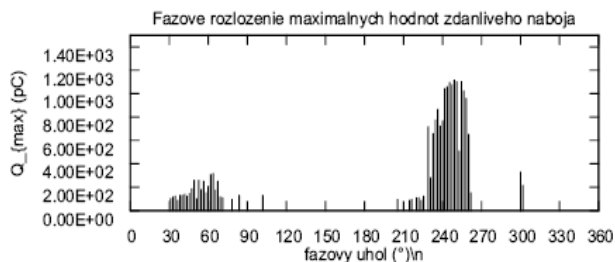
Fig. 6. Coil grounded through aluminum foil at 3.6 kV
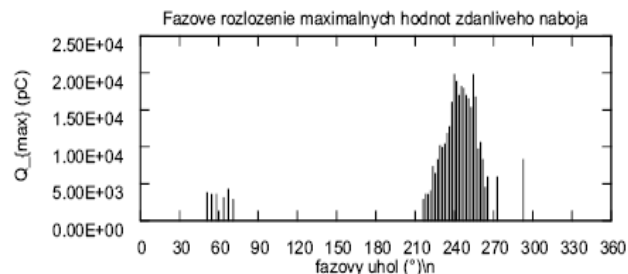
Fig. 7. Coil grounded through aluminum foil at 6 kV

C. Grounding of the slot part of the coil by four points

Fig. 8 and 9 show the phase distribution of maximum apparent charge for the voltage levels of 3.6 kV and 6 kV. Coil which was grounded by the four points showed discharge activity with due of q_{max} = 1000 pC (Fig. 10) at a voltage level of 3.6. The frequency of discharges rose to 0.08. It is probably the surface discharge at the end of the conductive layer.

The maximum value of apparent charge at a voltage level of 6kV was q_{max}= 19000 pC (Fig. 11) with a frequency of discharges of 0.8. These are surface discharges.

TABLE II
COIL GROUNDED BY FOUR POINTS

U [kV]	q_{max} [pC]	Nstr [-]	φ+ [°]	φ- [°]
3,6	1000	0,08	-	240-270
6	18000	0,8	20-80	220-300
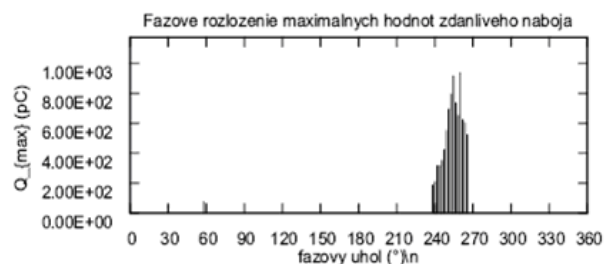
Fig. 8. Coil grounded by four points at 3.6 kV
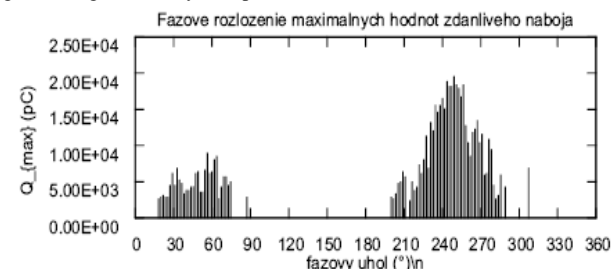
Fig. 9. Coil grounded by four points at 6 kV

IV. CONCLUSION

On the Fig. 10 is graphically illustrated the dependence of the maximum value of apparent charge by voltage. Based on measured results (Fig. 10) for 6kV voltage level, it can be said that the maximum apparent charge by grounding through a conductive mesh and using the aluminum foil reached 20000 pC. Grounding of the slot part of the coil by four points reached maximum apparent charge 19000 pC, the trend of increasing of the discharges activity was lower than at the other two types of ground (Fig. 12). Initial voltage for grounding through a conductive mesh was 3.4 kV, for grounding with aluminum foil was initial voltage 3kV and with the four points was initial voltage 3.6 kV. The advantages of measurement by the four points are higher levels of initial voltage, which affect the possible internal discharge activity in the studied coil and the lower value of the maximum apparent charge. When measured by the four points are not present in discharges, which occur between the foil respectively mesh and slot part of the coil. Finally this type of grounding reduced preparation time for measurement.
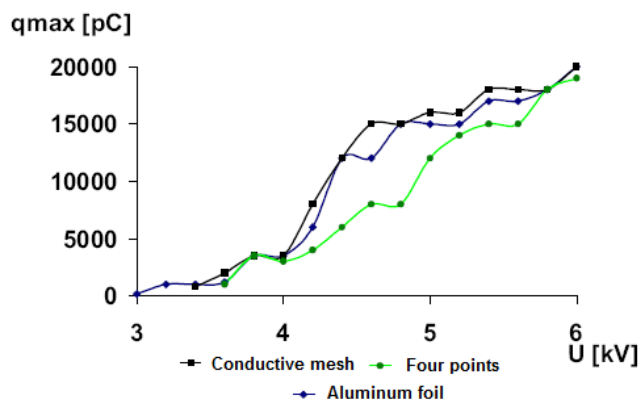
Fig. 10. Dependence of the maximum value of apparent charge by voltage

REFERENCES

[1] Kolcunová, I.: Diagnostika elektroenergetických zariadení metódou čiastkových výbojov, Košice, 2008, ISBN 978-80-553-0031-3.
[2] Kurimský, J,. Kolcunová, I., Cimbala, R.: Partial discharge analysis for insulation systems of electric rotating machines with various voltage stress. In: Acta Electrotechnica et Informatica. roč. 8, č. 4 (2008), s. 64-67.
[3] Cimbala, R.: Absorpčný prúd a jeho vzťah k iným diagnostickým metódam, Diagnostika 93, Plzeň 1993.
[4] E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," *IEEE Trans. Antennas Propagat.*, to be published.
[5] Petráš J., Džmura J., Balogh, J.: Metódy snímania čiastkových výbojov v izolačných systémoch vysokonapäťových zariadení, 2010. In: Electric Power Engineering 2010 : proceedings of the 11th international scientific conference : 4.- 6.5.2010, Hotel Santon, Brno, Czech Republic. - Brno : University of Technology, 2010. - ISBN 978-80-214-4094-4. - P. 707-710.

14

Measurement of Dielectric relaxation spectroscopy (DRS) on the Stator coils

Ľudovít Csányi, Martin Marci, Matúš Katin

Dept. of Electronics Power Engineering, FEI TU of Košice, Slovak Republic

ludovit.csanyi@tuke.sk, martin.marci@tuke.sk, matus.katin@tuke.sk

Abstract—**This paper discusses the mesurement of the stator coil with semiconductive protection and without semiconductive protection using the Dielectric relaxation spectroscopy. Effect of the electrode system on the charging current was comparison. The Electrode systems was simulate the position the coil in the groove of stator (2 point and 4 point electrode system). The measurement were made with DRS. DRS is non-destructive DC method.**

Keywords—**dielectric relaxation spectroscopy (DRS), rotating machine, insulation system, stator coil**

I. INTRODUCTION

The Growing demand of electricity means of the electric power engineering, that on the highvoltage electric equipment are asked claims of the safety and the reliability. Especially is the safety and reliability claims on the electrical insulation system, because he is one of the most important part of the electrical machines. By a [1] quarter of all failures are still related with the insulation systems.

On the Insulation have the greatest influence aging of the insulation, by [2] is to 31%. The Dielectric relaxation spetroscopy (DRS) is one of the method to determining state of aging of insulation system. The DRS is non – destructive method and he is on the base on a wide range measurment in the time domaine and frequency domaine [3].

II. INSULATION OF THE HIGH VOLTAGE ELECTRICAL ROTATING MACHINES

Rotating electric machines are used to convert electricity into mechanical energy and vice versa, depending on whether they work as a motor or a generator.

The asynchronous electrical machines are the most used electrical machines. Their power is from 100 W to hundreds kW. High-voltage insulation of stator force can be divided into [4]:

- insulation of coil and bars,
- spline section of isulation,
- insulation headboard of winding.

According to technology and mode of the binder we can divided insulation of high-voltage electrical machine to the two groups [5]:

Resin-rich technology

It is a three component composite (glass fabric, mica paper and suitable binder from 30 to 40 percent).

This composite is in a form of tape and is reeled
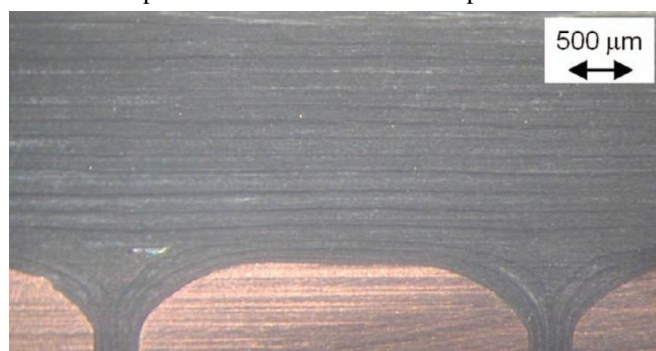
Fig. 1. Resin-rich structure [6].

continuously or discontinuously. The hardening takes place at a given pressure and temperature (160-170°C for 1 hour).

VPI technology

Primary role in this technology has absorbent mica tape, which is saturated by impregnant.
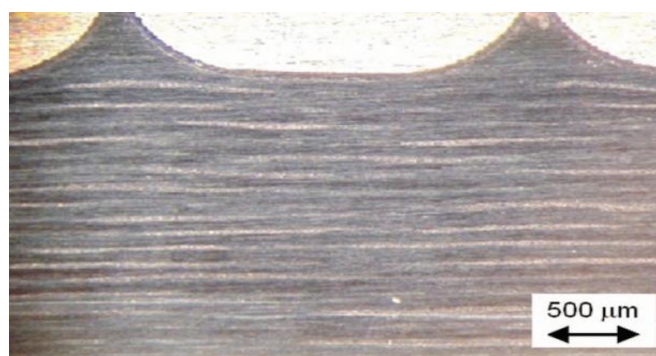
Fig. 2. VPI structure [6].

The VPI consists of three components (glass fabric, mica tape and binder 7 %).

Coils are multiwrapped by the tape. The impregnation process begins by drying at temperature 100°C for 20 hours. Object is vacuumed and flushed with glaze and then is drying and hardening using hot air.

III.   DIELECTRIC SPECTROSCOPY

For the diagnostic of dielectric materials was developed some of diagnostic methods. DRS is one of these methods.

It is a non-destructive method that is based on wide range of measurements in time and frequency domain and is included in area of Electrical Engineering, Physics, Chemistry and Material Engineering [7].

In general, properties of dielectrics can be characterized by the DC conductivity, relative permittivity and the dielectric response function.

Dielectric spectroscopy for the evaluation of the dielectric response in time domain use polarization and depolarization current and in the frequency domain use relative permittivity and loss number. DRS used these values for dielectric response evaluating.

IV.   POLARIZATION AND DEPOLARIZATION CURRENT

For Dielectric response measurement in time domaine, when a electric field $E(t)$ which is generate by an external voltage $U(t)$ is applied on homogenous dielectric material, the current through the test object can by expresed as[8]:

$$i(t) = C_0 \left[\frac{\sigma}{\varepsilon_0} U(t) + \varepsilon_r \frac{dU(t)}{dt} + \frac{d}{dt} \int_0^t f(t-\tau) U(\tau) d\tau \right]$$

(1)

Wher C_0 is the geometric capacitance of the test object, and σ is the conductivity. ε_0 and ε_r represent the permittivity of free space and the relative permittivity. $f(t)$ is the dielectric response function. The current $i(t)$ is the summation of displacement current and conduction current. When the test object is an arrangement of several dielectric materials in series or in parallel σ, ε_r and $f(t)$ represent the composite conductivity, relative permitivity and dielectric response function of this heterogenous test object, respectively [9, 10].

As show in Figre 3, assuming that a dc steps voltage $U(t)$ with the following characteristics is applied to a totally discharged test object.

$$U(t) = \begin{cases} 0 \_ t < 0 \\ U_0 \_ 0 \le t \le t_1 \\ 0 \_ t_1 < t \end{cases}$$

(2)

When $t < 0$, the current through the test object is zero and for time $0 \le t \le t_1$ the so called polarization current is generated due to the conductivity of the test object and the various polarization processes within the test object. The Polarization current can by written as [8]:

$$i_p(t) = C_0 U_0 \left[\frac{\sigma}{\varepsilon_0} + f(t) \right]$$

(3)

At time $t = t_1$ the external voltage is removed and the test object is short-circuited. According to the superposition principle, the depolarization current can by expressed as [8]:

$$i_d(t) = -C_0 U_0 \left[f(t) - f(t + T_c) \right]$$

(4)
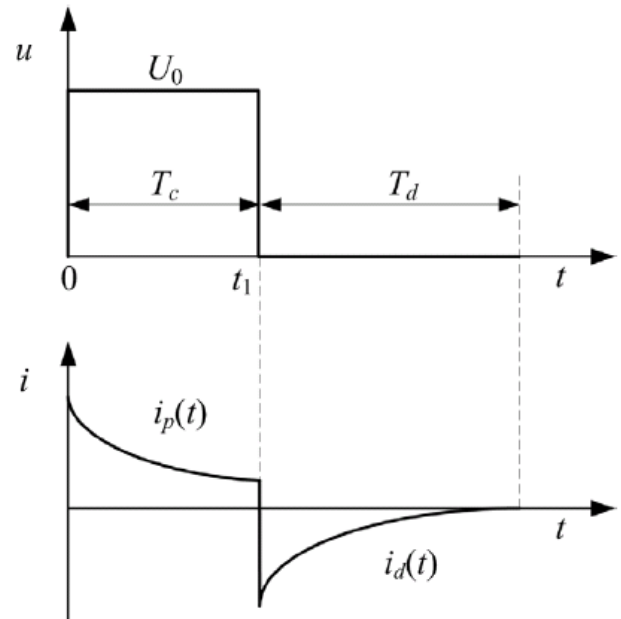
Fig. 3.   Principle of polarization and depolarization currents measurement [8].

T_c which is from $t = 0$ to t_1 represent the polarization duration time and T_d represent the depolarization duration time.

V.   EXPERIMENT

The object of measurement was comparison of two electrode systems with semiconductive protection and without semiconductive protection. All measurements was 12, because 6 measurement was on coil with semicinductive protection (3 measurements on 2 point electrode system and 3 measurement on 4 point electrode system) and other 6 measurement was on coil without semicodnuctive protection (the same number as for the semiconductive protection).

The scheme of measurement is the same for both electrode systems (Fig.4).
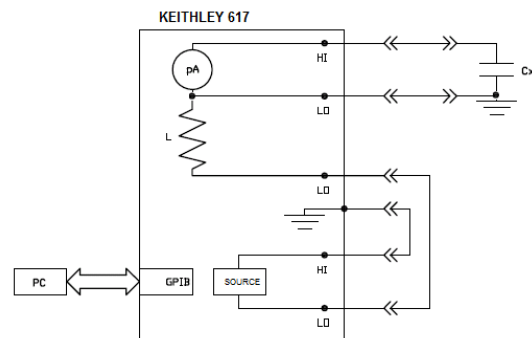
Fig. 4. Scheme of measurement.

Description of involment:

- C_X –measurements coil,
- U – elekctrometer KEITHLEY 617,
- GPIB – bus,
- PC – computer.

Two point electrode system was measurement first (Obr.5), for the coil with semiconductive protection and for the coil withou semiconductive protection. Two piont electrode system was created from two conducting electrodes wound in the middle of the coil width 10cm ($\pm$0,2cm). These electrodes was short-circuited and involved into the measurement circuit.

The current dependence on time we see on the next graph (Fig.6-11) for the two point electrode system.
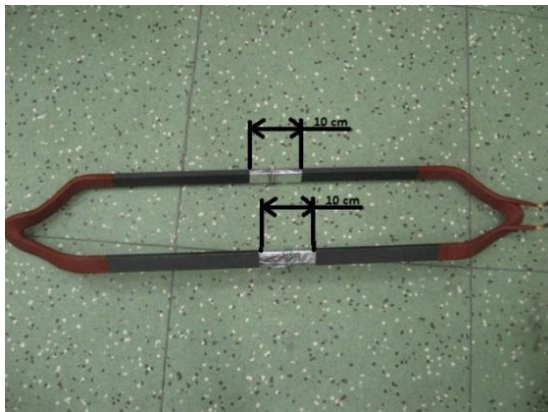
Fig. 5. Two point electrode system
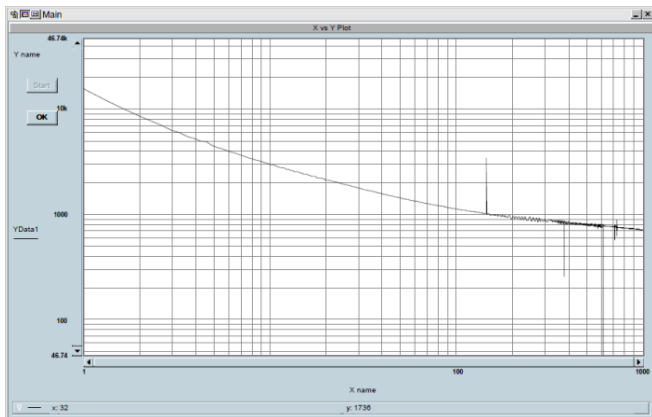
Fig. 6 The Dependence of coil no.1 with semiconductive protection (2 point electrode system)
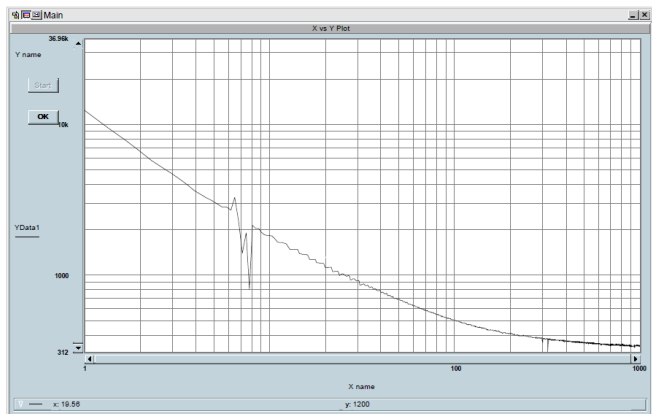
Fig. 7 The Dependence of coil no.1 withou semiconductive protection (2 point electrode system)

Fig. 8 The Dependence of coil no.2 with semiconductive protection(2 point electrode system)
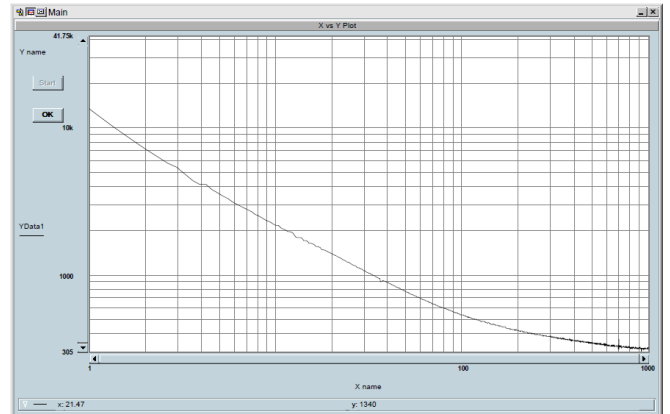
Fig. 9 The Dependence of coil no.2 without semiconductive protection (2 point electrode sysem)
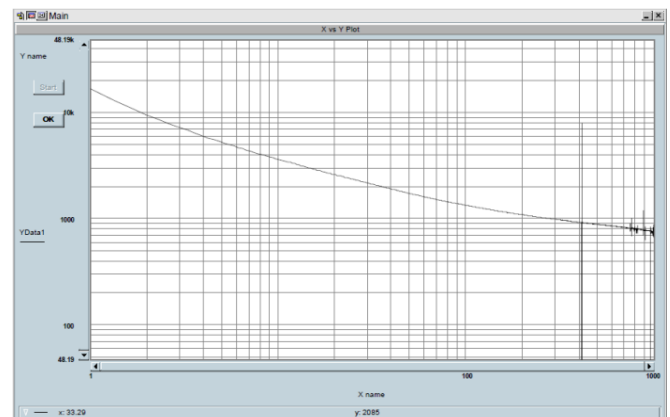
Fig. 10 The dependence of coil no.3 with semiconductive protection (2 point electrode system)
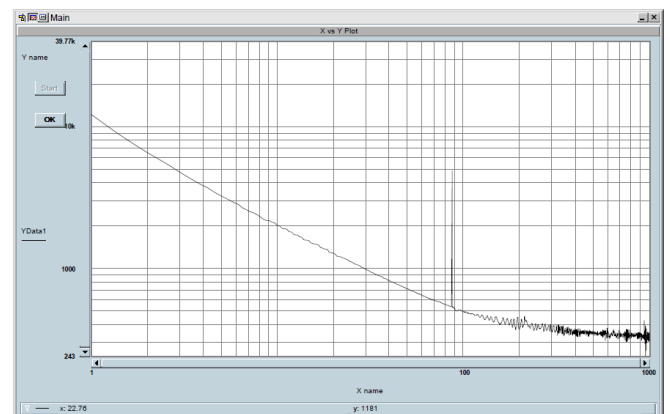
Fig. 11 The Dependence of coil no.3 without semiconductive protection (2 point electrode system)

14

We continue with the four point electrode system (Fig.12). The Electrode system of four point being four electrode wound on end of spline section of coil. Width of electrode is 5cm ($\pm 0,2$cm).

We see on the next graphs (Fig.13-18) the current dependens on the time, for the four electrode system.
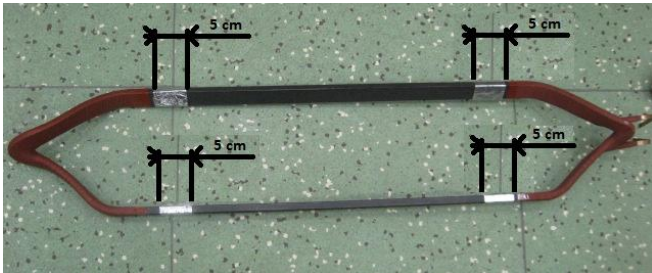
Fig. 12 The Four point electrode system

Fig. 13 The Dependence of coil no.1 with semiconductive protection (4 point electrode system)
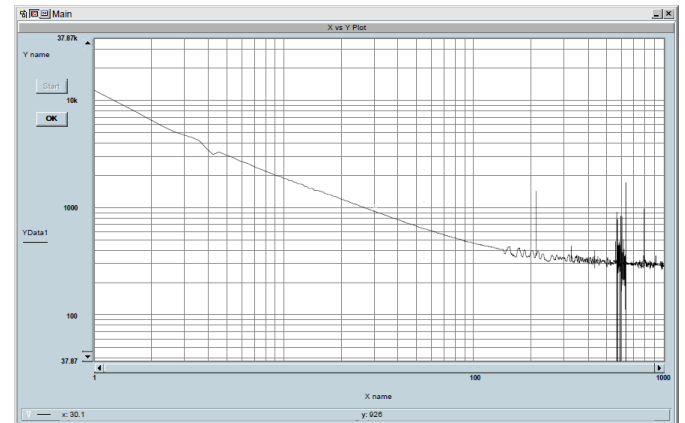
Fig. 14 The Dependence of coil no.1 without semiconductive protection (4 point electrode system)
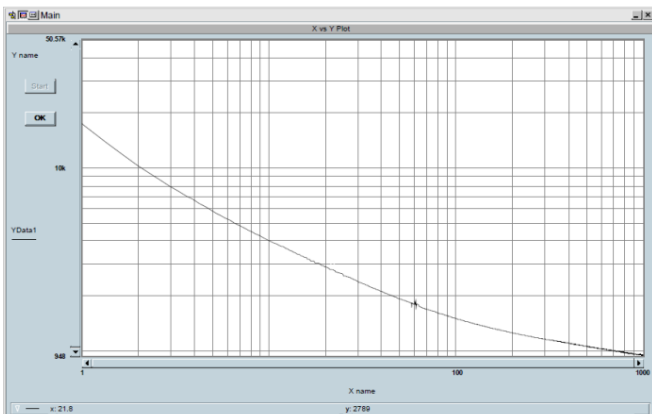
Fig. 15 The Dependence of coil no.2 with semiconductive protection (4 point electrode system)

Fig. 16 The dependence of coil no.2 without semiconductive protection (4 point electrode system)

Fig. 17 The Dependence of coil no.3 with semiconductive protection (4 point electrode system)

Fig. 18 The Dependence of coil no.3 without semiconductive protection (4 point electrode system)

The Next tables (Tab.1-6) show clearing comparison of dependence. These tables represent the current dependence for the two point and four point electode systems. We see in the tables time konstant and elementary current.

Tab.1 Seven aproximation of coil no.1 with semiconductive protection

7 approximation coil no.1			
2 points		4 points	
τ_i (s)	I_{mi} (pA)	τ_i (s)	I_{mi} (pA)
0.14	306027	0.16	272650
0.56	34022	0.64	27822
2.20	8140	2.67	7728
7.95	2182	14.13	2382
27.96	1353	61.01	818
135.38	499	527.08	411
5145.07	870	45302.45	779

Tab.2 Seven aproximation of coil no.2 with semiconductive protection

7 approximation coil No.2			
2 points		4 points	
τ_i (s)	I_{mi} (pA)	τ_i (s)	I_{mi} (pA)
0.18	184587	0.17	240107
0.73	26414	0.74	24797
3.09	6832	2.97	7231
14.21	2687	13.08	2693
56.97	1170	47.17	1274
305.98	448	237.48	451
19329.82	811	7343.20	1080

Tab.3 Seven aproximation of coil no.3 with semiconductive protection

7 approximation coil No.3			
2 points		4 points	
τ_i (s)	I_{mi} (pA)	τ_i (s)	I_{mi} (pA)
0.15	263021	0.16	236960
0.67	30366	0.66	29830
3.09	7161	2.98	7349
15.65	2377	14.24	2015
60.61	901	66.62	490
183.57	355	602.70	427
3733.35	993	492835.83	798

Tab.4 Seven aproximation of coil no.1 wihout semiconductive protection

7 approximation coil no.1			
2 points		4 points	
τ_i (s)	I_{mi} (pA)	τ_i (s)	I_{mi} (pA)
0.13	264686	0.16	189083
0.55	27013	0.69	21137
2.39	8565	2.67	6521
15.66	1583	14.30	1569
77.10	433	70.91	453
721.60	74	603.43	79
223305.13	321	265438.55	374

Tab.5 Seven aproximation of coil no.2 wihout semiconductive protection

7 approximation coil no.2			
2 points		4 points	
τ_i (s)	I_{mi} (pA)	τ_i (s)	I_{mi} (pA)
0.18	178128	0.16	213311
0.73	21306	0.62	28804
2.92	6100	2.48	7456
14.57	1950	10.85	2365
52.86	617	38.42	955
308.45	185	211.79	223
1090906.64	310	1203291.63	483

Tab.6 Seven aproximation of coil no.3 wihout semiconductive protection

7 approximation coil no.3			
2 points		4 points	
τ_i (s)	I_{mi} (pA)	τ_i (s)	I_{mi} (pA)
0.16	187362	0.10	446756
0.54	25017	0.41	38590
2.14	7405	1.59	10570
10.47	2328	7.43	2404
53.53	733	32.38	1038
325.78	107	195.00	209
1676182.29	328	209529.38	294

14

VI. Conclusion

The Results of this paper is comparison of two electrode systems on 6 kV coil with semiconductive an and without semiconductive protection. The Result were obtained by dielectric relaxation spectroscopy (DRS).

The time constant and elementary current were obtained with measurement results. All measurement result are in the time domaine. The transform from time domaine into the frquecy domaine will by subject of the further researche.

Acknowledgment

References

[1] THORSEN, O.V. DALVA, M.: "A survey of faults on induction motors in offshore oil industry, petrochemical industry, gas terminals and oil refineries, „ IEEE Trans. Ind. Appl. vol. 31, no. 5, pp. 1186-1196, Sep./Oct. 1995

[2] CIGRE Study Committee SC11, EG11.02, "Hydrogenerator Failures – Results of the Survey, " 2003

[3] Farahani, M., Borsi, H., Gockenbach, E. Dielectric spectroscopy in Time and Frequency Domain on Insulation System of High Voltage Rotating Machines. In *8th IEEE International Conference on Solid Dielectrics (ICSD)*. Toulouse, France, 2004.

[4] MÁLEK, J., KOUBÍK, L.: Slídové izolanty pro vysokonapěťové elektrické stroje. In: Diagnostika 99, Plzeň: ZČU 2003.

[5] PETR, J.: Izolačný systémy elektrických strojů, FEL CVUT, Praha, Dostupné na internete: http://martin.feld.cvut.cz/~petr/Izolsyst.pdf.,

[6] VOGELSANG, R WEIERS, T., FRÖHLICH, K., BRÜTSCH, R., Electrical Breakdown in High-Voltage Winding Insulations of Different Manufacturing Qualities, IEEE, Electrical insulation magazine, vol. 22, no. 3, May/June 2006.

[7] Palai-Dany, T.: Dielektrická spektroskopie karboxymetylcelulózy v časovej oblasti, Zkrácená verze Ph.D. Thesis, 2009, CVUT Brno, FEKT, Ústav fyziky.

[8] SHI-QIANG, W. GUAN-JUN, Z. JIAN-LIN, W. [14] SHUANG-SUO, Y. MING, D. XIN-BO, H.: "Investigation on Dieletric Response CHaracteristics of Thermally Aged Insulating Pressboard in Vacuum and Oil-impregnated Ambient", IEEE Trans. Ind. on Diel. and El. In., vol. 17, No.6, pp. 1853-1892, December 2010.

[9] W.S. Zaengl, „Dielectric spectroscopy in time and frequency domain for HV power equipment, Part I: theoretical consideration", IEEE Electr. Insul. Mag., Vol. 19, No. 5, pp. 9-22, 2003.

[10] A. K. Jonscher, Dielectric relaxation in Solids, London, UK, CHealsea dielectrics Press, pp. 42-47, 1983.

Measuring of corona discharge in liquid dielectrics

[1]Martin MARCI, [2]Ľudovít CSÁNYI

[1, 2] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]martin.marci@tuke.sk, [2]ludovit.csanyi@tuke.sk

Abstract—**Insulation is one of the most important part of every electrotechnical equipment. There is a big constrain to develop new, more ecologically friendly electrotechnical materials for this purposes. There is one new trend in the liquid dielectrics field, concrete developing of electroinsulating oils based on vegetable base. This paper deals with comparison of silicone oil sample and colza oil sample due to resistance against corona..**

Keywords—**Liquid insulant, vegetable oil, mineral oil, corona discharge.**

I. Introduction

For the correct function of every power device, high voltage insulation has to resists the operation conditions. Due to the high values of operating voltage and the occasional extreme conditions, in the insulation of power devices a partial discharge can occur in the areas, where the inhomogeneity degree of the electric field is too high. This discharge activity degrades insulation, creates the gas bubbles, the solid impurities and changes the chemical composition of the environs. This process could lead to gradual degradation, or even to the total destruction of insulation. In the laboratorial conditions it is possible to investigate this discharge using the electrode system point electrode – disk electrode, dipped into the insulating oil. This is the way how to investigate an influence of pollution of liquid dielectrics as well as an influence of the sharp shoulders, electrode borders and the other construction elements at the electric strength of various liquid dielectrics.

The partial discharge, in the air known as corona, is referred to as streamer in the liquid dielectrics. Intensity of the discharge activity is closely associated with an ageing and damaging of the insulation. Because of that, it is necessary to understand to this process in order to be able to define a condition of the insulation.

Corona discharge is discharge in heavy inhomogeneous electric field, which is caused by ionization of the liquid dielectric ambient a conductor. It occurs in the moment, when the potential gradient in the inhomogeneity area overreach a specific value, but electric strength of remaining part of the insulation is sufficiently high to resist the full breakdown of the whole insulation.

Corona is process by which ions generated by the electric field cross from the areas with higher potential into the near areas with lower potential, or recombine and create neutral molecule. When potential gradient in the liquid dielectric is equally high, environs in this area become ionized and subsequently conductive. In strong electrical fields, conductivity of the liquid dielectrics depends on purity level, shape of the applied voltage, electrode system configuration, quality status of the surface of the electrodes and used material of the electrodes. In the very pure liquids, a dependence of electric field intensity on the current abounding in the interelectrode space has three areas. First area corresponds to the intensity value less than $10 kVcm^{-1}$. In this area, a particles migration is affected by the basic Ohms law and the major task belongs to the ion and cataphoretic conductivity. The second area, in which the current shows just a little dependence on the electrical field intensity is called zone of saturation and corresponds to the intensity from $10 kVcm^{-1}$ to $100\ kVcm^{-1}$. In the last area where the electric field intensity is higher than $100\ kVcm^{-1}$, increasing of conductivity of the liquid causes the high dependence of abounding current on the increasing values of the electric field intensity.

Increasing of the liquid dielectric conductivity, could be caused by increasing of the charge carrier mobility. This is incured by increasing of voltage value and by enhancement of charge carrier concentration. Enhancement of charge carrier concentration in the electric fields with high intensity value may be brought on by the additional creation of ions following intensive dissociation of the molecules of liquid and foreign substance by a cold emission (emission by field), or by a thermo-ionic emission (schottky effect), which is facilitated by strong electric field – modification of a barrier, which the electrons have to overcome at escape from the metal, and by collision ionization in the liquid capacity.

II. Experiment

A. Measured sample and electrode configuration

Point – disk electrode configuration was used for corona measuring. Corona was investigated in the two oil samples, concretely in silicone oil - Lukosiol M200 with viscosity of $200 mm^2 s^{-1}$ on 25°C and in vegetable colza oil with viscosity of $62 mm^2 s^{-1}$ on 25°C. Discharge activity was created on a spike of the point electrode, where the AC high voltage was applied.

B. Measuring workspace arrangement

Concrete partial discharge measuring was made by direct method with serial connection of measuring impedance and

coupling capacitor as shown on Fig. 2. The principle of this method is described in IEC 60270 standard.
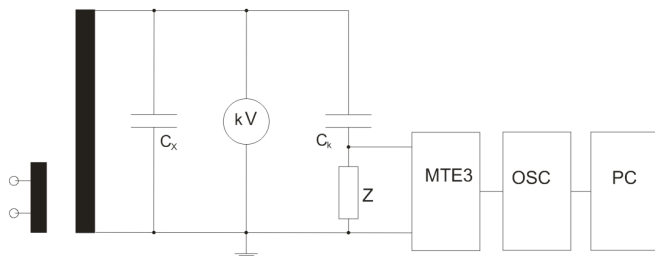
Fig. 1. Diagram of partial discharges measuring circuit [1]

Partial discharges emergent on measured object C_x in the form of current impulses are transformed to voltage impulses with the help of measuring impedance Z. These are processed by the measuring device **MTE3**, which is able to receive and display partial discharges with various amplitudes in dependence on setting the value of sensing sensibility and allocate them to corresponding phase angle. Digital oscilloscope **OSC** saves scanned impulses into memory and works like an **A/D** converter. In the next step, data are sent into a personal computer, where they are compiled upon convenient calibration by a program (created in HP VEE environment). Its aim is to collect data and save them on a storage medium in an appropriate manner. Later, data are processed into the final form.

At last, measuring outputs were compiled into the φ, q, n graphs. Measured object C_X consists from glass bank filled by measured sample of the liquid dielectric, the high voltage and grounded electrode (Fig. 2).
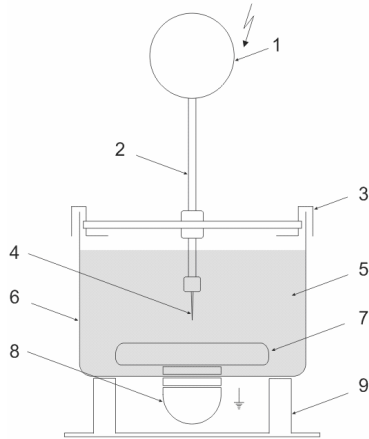
Fig. 2. Measured object
1 – homogenizing electrode of the high voltage electrode feeding, 2 – supply of the high voltage electrode and a screw to set an interelectrode distance, 3 – fixative staples, 4 – high voltage electrode, 5 – sample of measured dielectrics, 6 – glass tank, 7 – disk electrode on the earth potential, 8 - homogenizing electrode of the electrode on the earth potential, 9 – insulating pad

C. Measuring procedure

Measured sample of liquid dielectric were poured into the glass bank so that the smallest quantity of air bubbles penetrate into the oil. After connecting of the measured object into the measuring circuit, there was setting interelectrode distance between point electrode and disk electrode in 35mm. There was applied the AC high voltage on the high voltage electrode, which was increased until first discharge occurs. After the discharge measuring for concrete voltage level,

voltage value was increased by 1kV. Measuring was made for every voltage value up to 30kV. Measuring was made for interelectrode distances d = 20mm and 30mm too. Due to the large number of gained results, this work does not deal with these distances.

III. EVALUATION OF THE MEASUREMENT RESULTS

The measuring results show different behavior of a creation and a progress of corona discharge in the air and in the liquid dielectric. In the liquid dielectric, ignition discharge occur in a positive half-wave of the applied harmonic voltage behind 90° following the phase (in the area of 110°) (Fig. 3a). Increasing of applied voltage causes occurrence of discharge in a negative half-wave behind voltage amplitude (in the area of 290°) (Fig. 3b). Discharge amplitude is bigger in negative half – wave.
With the increasing of the applied voltage, discharge activity expands following the phase in the direct of U_{max}. Intensity of discharge activity is very unstable and easily become extinct at lower voltage levels.
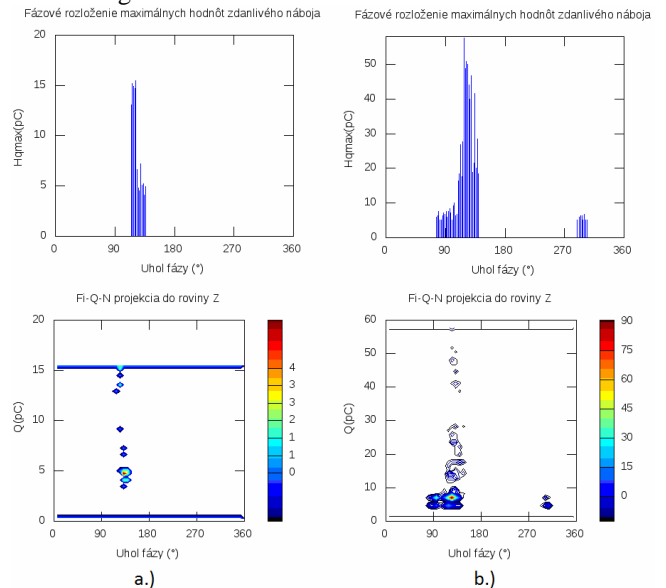
Fig. 3. a) φ, Q,N of discharge activity in Lukosiol, 7kV
b.) φ, Q,N of discharge activity in Lukosiol, 9kV
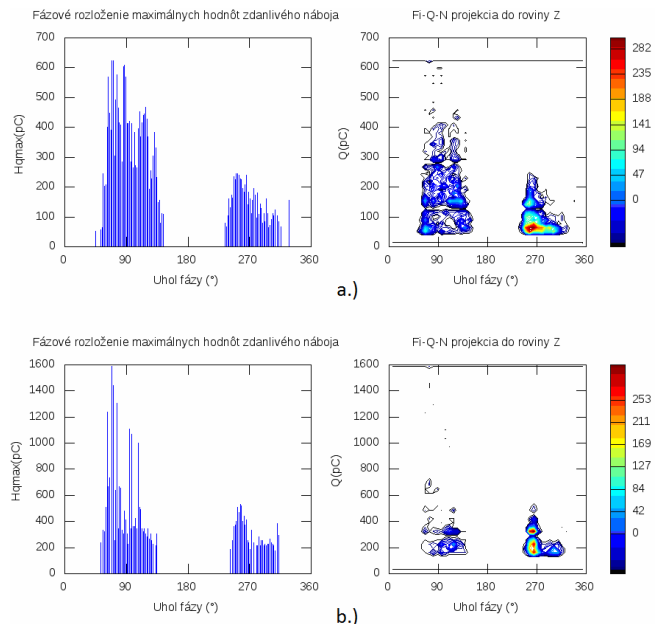
Fig. 4. a) φ, Q,N of discharge activity in Lukosiol, 26kV
b) φ, Q,N of discharge activity in Lukosiol, 28kV

Fig. 5. a) φ, Q,N of discharge activity in colza oil, 26kV
b) φ, Q,N of discharge activity in colza oil, 28kV

TABLE I
EVALUATION OF THE PARAMETERS OF THE DISCHARGE ACTIVITY

Applied voltage	Q_{max} [pC] (colza oil)		Q_{max} [pC] (Lukosiol)	
	+	-	+	-
26 kV	500	300	620	250
28 kV	680	480	1600	600

From the comparison of discharge activity with increasing of applied voltage in silicone oil (obr.4 - a, b) and vegetable oil (obr.5 - a, b) it is able to observe some conclusions:

- progress of discharge activity is similar in both samples – with voltage increasing, discharge activity grows and has the same phase arrangement.
- stability of discharge activity strongly depends on viscosity of the liquid dielectric,
- ignition voltage of corona discharge creation is higher for vegetable oil sample (12kV) than in silicone oil (7kV),
- maximum value of image charge of partial discharge for the same voltage levels is higher for silicone oil (Fig. 4,5), (Table 1).

IV. CONCLUSION

This paper deals with the research of corona discharge activity development in the liquid dielectric and with comparison of samples of silicone and vegetable oils. From the view of corona discharge activity development, vegetable oil showed very good results. Because of that, it is necessary to continue with measuring of next samples of liquid dielectrics and to determine influence of humidity and temperature on development of corona discharge activity.

ACKNOWLEDGMENT

REFERENCES

[1] J. Tkáč, J. Kurimský, B. Dolník *"Laborátorný výskum výbojovej činnosti v modeloch s kombinovanou izoláciou olej – papier,"* In: Starnutie elektroizolačných systémov, 4/2008. Dostupné na internete: < http://web.tuke.sk/fei-kee/jses/uploads/File/jses-04-2008-07.pdf>

[2] Suwarno *"Phase-resolved analysis of discharges in air, liquid and solid insulating materials,"* Power Engineering Conference, 2005. IPEC 2005. The 7th International, Nov. 29 2005-Dec. 2 2005.

[3] R. Cimbala *"Tepelná závislosť izolačného stavu rastlinného oleja,"* In: Starnutie elektroizolačných systémov, no. 3 (2007), Dostupné na internete: < http://jeen.fei.tuke.sk/index.php/JSES/article/viewFile/101/103>

[4] J. Džmura, J. Pettáš, J. Balogh *"Partial discharge measurements using acoustic mehods,"* In: Electrotehnica, Electronica, Automatica. vol. 56, no. 4 (2008), p. 10-12. ISSN 1582-5175.

[5] J. Tkáč, J. Kurimský, B. Dolník *"Laborátorný výskum výbojovej činnosti v modeloch s kombinovanou izoláciou olej - papier,"* Publikácia TUKE: Starnutie elektroizolačných systémov, 4/2008. Dostupné na internete: < http://web.tuke.sk/fei-kee/jses/uploads/File/jses-04-2008-07.pdf>

[6] M. Marci *"Meranie výbojovej činnosti v kombinovanej izolácii olej – papier,"* diplomová práca, Košice, 2009.

Measuring partial discharges on coils without corona protection

[1]Marián Hrinko, [2]Pavol Hocko [3]Lýdia Dedinská

[1]Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic
[2]Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic
[3]Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic

[1]marian.hrinko@tuke.sk, [2]pavol.hocko@tuke.sk, [3]lydia.dedinska@tuke.sk

Abstract— with a directive growth of development of new insulating materials and trends, which tend to minimalization of electrical rotation machines, we have to hold in mind verifying the quality of these materials. The insulations are exposed to higher field strength by decreasing of its thickness. We could say that reciprocal amount of electrical stress corresponds to voltage life what means decreasing the voltage life in final merit. Appropriate method for verifying the quality of insulation systems is measurement of partial discharge, which was used in this case.

Keywords — Stator insulation, stator coil, measurement partial discharges.

I. INTRODUCTION

During the operation of machine, the insulation is subjected to combination of degradation factors. That cause different changes in internal structure of insulation material which could be expressed by dropping quality of insulation system in outwards merit. Seeing that the quality is close-knit with reliability and no-failure operation, there is the necessary condition these factors eliminate. For this purpose there is a diagnostic of electrical rotation machines that detect state of such machines and which target is to ensure the longest no-failure operation machines [3].

One of many diagnostic methods for testing insulation state of rotating machines is partial discharges measurement. The outputs from this method are variety parameters that help us better know inner state of insulation and basically on these information determine as the quality of insulation is satisfied. For the full analysis of parameters is being demanded extra factor, which cannot be measured – human experience. Accordingly without the measured parameters is important the human eye too. In the now-days is being possible alternative of human experience – the neuron network. But for well processing and evaluating purposes the neuron network must be taught. The input parameters to the neuron network for teaching process can be used parameters from the measurements of this article.

II. MEASURING SAMPLES AND MEASURING

For testing the origin of partial discharges were used four high voltages 6 kV stator coils with new wrapped insulation (measuring samples). The coils was built up for asynchronous machine with the rated phase-to-earth voltage Un = 3,6 kV, the rated current In = 27,6 A and the rated power Pn = 200 kW. Main insulation of samples was Samicatherm 366.28 based upon a woven glass fabric, calcined mica paper and a thermosetting epoxy resin as binder. The semi-conductive protection was CoronaShield P800 – a modified phenolic resin with semi-conductive filler.
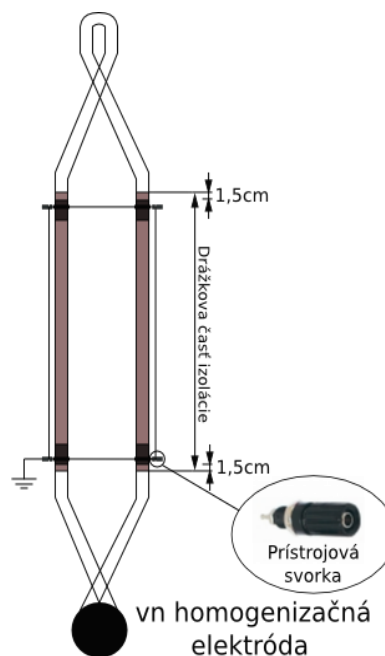
Fig.1 The earth connection of slot part of coil.

The coil was hung on the insulation rope placed in Faraday cage. End of the coil was terminated by homogenous electrode for preventing corona discharges and connected with high voltage potential. The slot part of the coil was wrapped with conductive foil 5 cm width in the specific places (1,5 cm from ends of slot). Individual foils was together galvanic connected

and driven on potential earth. This setup models putting coil in the stator slot.

III. EXPERIMENTAL SETUP

The straight detection of partial discharges was used for measurement. Measuring impedance was placed in the series with the coupling capacitor Cv. The great advantage of straight detection is no destruction of measurement equipments if measuring sample breaks down.

All measurement system consists of three parts:
1. High-voltage part: HV transformer, coupling capacitor C_v, sample C_x,
2. Measurement part: measuring impedance Z_m, equipment for detection of partial discharges *MTE3,*
3. Assessment part: digital oscilloscope with internal memory, personal computer with software for analysis of partial discharges.
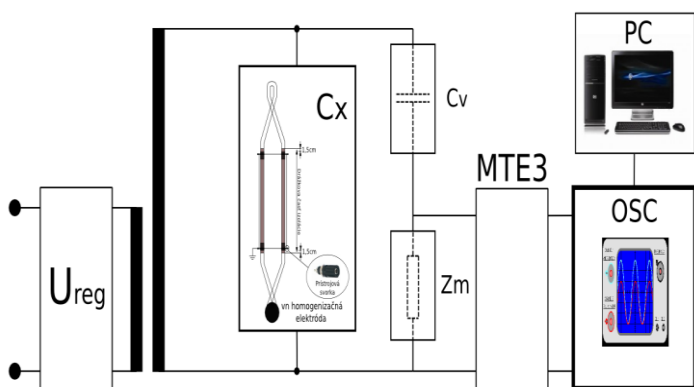
Fig.2 Scheme of experimental setup.

The experimental setup is sketched on Fig 2. U_{reg} is controllable voltage source used for controlling voltage on the sample. C_v is coupling capacitor that supply charge to circuit after voltage drop caused by discharges in sample C_x. *MTE3* is the measuring equipment used for entrapping and enhancing voltage drops on the measuring impedance Z_m caused by discharge activity. The entrapped and enhanced discharges continue to digital oscilloscope with memory, where it is transformed to binary code. This code is workable to computer. Computer by appropriate application execute entrapped and enhanced discharges in binary code and store it into the file. Then this file with measured values of partial discharges is processed to simple plots that we used for deducting useful values.

IV. MEASURING PROCESS

The first setup voltage was 2kV with time duration three minutes. If stabile partial discharges were not entrapped during setup time the voltage increased with the same time duration. It was repeated if no stabile partial discharges appeared (The Inception Voltage U_i). After this the determinate program for measurement of partial discharges was launched and it was noting and was evaluating the entrapped partial discharges. Next measurements were

performed with step 200 V to 6 kV. During time duration was noted more than 900 periods of applied voltages.

V. MEASURED VALUES

For purpose evaluating and comparing individual coils was set three voltage values:
- The inception voltage of partial discharges U_i,
- The rated phase-to-earth voltage of partial discharges $U_{npe} = 3{,}6$ kV,
- The rated phase-to-phase voltage of partial discharges $U_{npp} = 6$ kV.

At the each voltage values were registered:
- The crest value of the image charge q_{max},
- The multitude of partial discharges n_{str},
- The phase angle of origin of the partial discharges at the positive and at the negative half-period φ+/φ-.

.

TABLE I
MEASURED VALUES FOR COILS

U [kV]	q_{max} [pC]	n_{str} [-]	φ+ [°]	φ- [°]
Coil No.1				
3,8	160	0,2	0-130	180-330
6	1100	1	0-150	170-360
Coil No.2				
3	110	0,15	0-130	180-330
3,6	200	0,15	0-110	180-330
6	3000	0,15	0-150	180-330
Coil No.3				
3,8	230	0,05	10-110	10-270
6	2500	0,6	0-150	180-330
Coil No.4				
3,2	23	0,1	0-80	220-270
3,6	55	0,1	0-120	180-300
6	4800	0,1	10-150	180-330

Measuring values of image charge, multitude and phase angle of origin of the partial discharges for measured coils are shown in Table 1 above.

VI. DISCUSSION ABOUT MEASURED VALUES

Measured multitude values of image charge for individual coils are sketched on Fig.3. The maximal value of image charge reached coil No.2 concrete 5000 pC at the voltage values 5,6 kV and 5,8 kV. Then fell down to 3000 pC at the voltage value 6 kV. This coil disposed by the lowest value of

the inception voltage $U_i = 3$ kV with crest value of the image charge 110 pC. The second coil in term of the inception voltage was coil No.4. The characteristic of the crest values of image charge was steeper than in the coil No.4, the higher values was denoted at the voltage values 5,4 kV and 6 kV. The inception voltage of image charge was 3,8 kV what is 800V more than in previous case. The trend of image charge for coil No.3 is resembled with the trend of coil No.4 to the voltage value 5,2 kV. The crest value of image charge reached 4000 pC at the voltage value 5,8 kV. Ultimately coils No.2, No.3 and No.4 showed resembling discharge activity. The best values of image charge showed coil No.1. Stabile discharge activity was entrapped at the voltage value 3,8 kV what is the same value as in coil No.4, however the crest value of image charge did not exceed 1100 pC what is only one fourth of values of the image charge other coils.
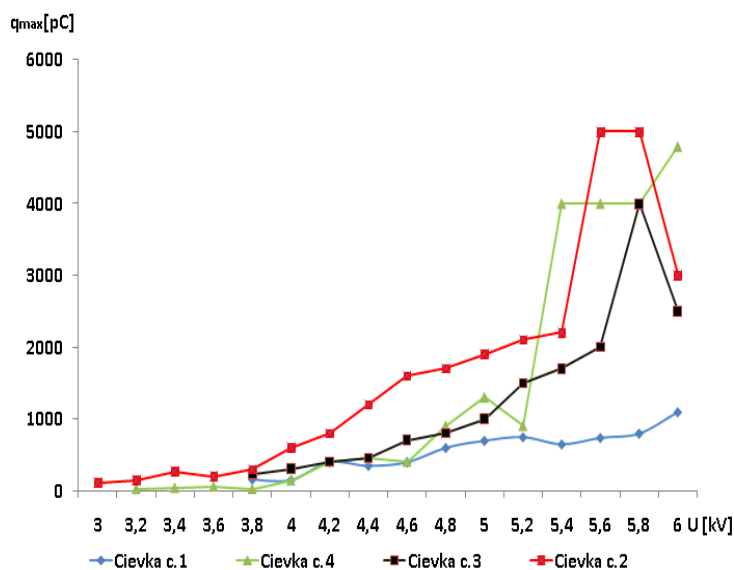
Fig.3 . Waveforms of the image charges of individual coils No.1 – No.4.

VII. Conclusion

It can be proclaimed that the inception voltage indicate the quality of insulation. Lower amounts can point to worse quality (more cavities in insulation), higher amounts can point to better quality (less cavities in insulation). The presence of cavities can be caused by manufacturing process. The higher values of image charge can touch the lower quality of application of corona protection in the end of slot part of coil. Basically we could classify the coils, concrete coils No.2 and [14] No.3 contain more cavities and duties in insulation and they have worse quality of application of corona protection. The coils No.1 and No.4 contain less cavities and duties, but application of corona protection is better for coil No.1.

Using of corona protection at the end of the slot part of coil decrease discharge activity to available amounts – in the worst case the crest value of image charge did not exceed the 5000 pC. So we can say that corona protection has positive influence to discharge activity.

References

[1] G VEVERKA, Antonín: Technika vysokých napätí, SNTL/ALFA 1982, 04-527-82, 304 strán, 164 obrázkov, 23 tabuliek.

[2] KOLCUNOVÁ, I,. KVAKOVSKÝ, M.,KURIMSKÝ, J.: Meranie povrchových výbojov na vn cievke točivého stroja. In: Elektroenergetika Journal. 4 (2009), ISSN 1337-6756

[3] MARTON, K., : Technika vysokých napätí 1, ALFA, Bratislava 1983

[4] KOLCUNOVÁ, Iraida: Diagnostika elektrických strojov, Technická univerzita Košice 2006, ISBN 80-8073-550-6.

Mobile minimachine for inspection of constrained spaces

[1]Pavol CABÚK, [1]Michal JURČIŠIN, [2]Peter KAŤUCH

[1]Dept. of Technologies in Electronics, FEI TU of Kosice, Slovak Republic
[2]Department of Biomedical Engineering, Automation and Measurement,
Faculty of Mechanical Engineering, TU of Kosice, Slovak Republic

pavol.cabuk@tuke.sk, michal.jurcisin@tuke.sk, peter.katuch@tuke.sk

Abstract—**Many research centers over the world solved problems with mobile robots for rough terrain and for human inaccessible places. There are more advantages of described solution. Miniature robot is able to work in constrained compartments and with installed camera system provide look on around environment. Added value is possibility of chassis replacement to optimal and fast replacement of contact elements (whegs, wheel, belts). In final state the designed control electronics will be based on heterostructure LTCC structure - HTCC substrate with shaped part for sensor placement.**

Keywords— **miniature robot, heterostructure, legs**

I. INTRODUCTION

Mobile machines designed for moving in varied terrains frequently uses complicated chassis for walking or wheel driving. Miniaturization of such types of chassis frequently collides with technological limits, which not allow chassis miniaturization to required size. Other reasons are time and economical requirements. Searching of simple and smart solution provide possibility of shift those limits. Developed mobile minimachine is designed for pipe inspection systems with capability for work in other constrained compartments. Required functions are: moving forward, moving reverse, turning round with minimal radius, detection and obstacles avoidance, in future with wireless CCD camera transmission.

II. MECHANICAL DESIGN

For construction concept for mobile minimachine was in cooperation with Faculty of Mechanical Engineering TU of Kosice used, construction of chassis allows work with wheels, straps or legged wheels, which joins advantages of wheels and legs. A legged wheel offers reliable and fast way for move in various terrains in combination with simple construction. Main advantage of wheels is higher speed limit. Main advantage of legs is easily negotiation of terrain. Combination of these two advantages is new type of component shown on Fig. 1. Inspirations for use of these components are Whegs robots developed by Biologically Inspired Robotics Laboratory on Case Western Reverse University
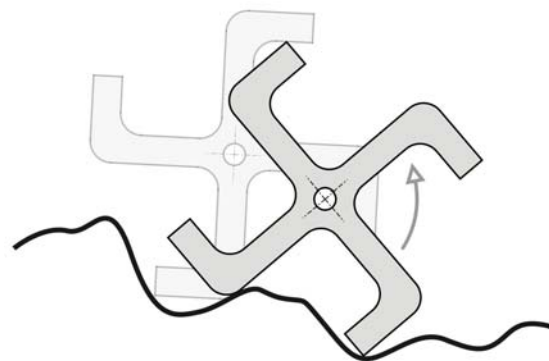
Fig. 1. Legged wheel (wheg)

Mainframe (Fig. 2) was milled from mono-block of PMMA material (polymetylmetacrylate).
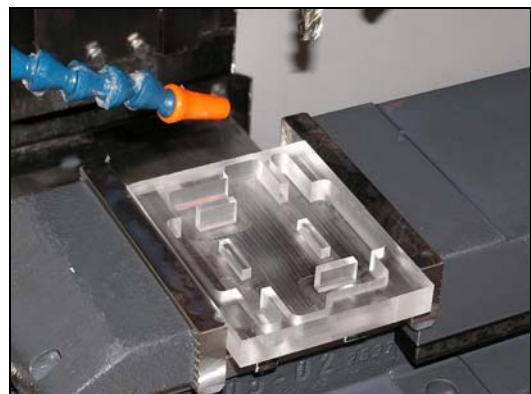
Fig. 2. Mainframe milled from mono-block of PMMA

Construction of mainframe is typical for simplest caterpillar chassis. Used conception of four wheel chassis worked with two divided axles. Left and right side of chassis are independently powered by DC motors with gearboxes. Connection with front and rear propulsive components is realized by belt transmission. On Fig. 3 is schematically shown propulsion subsystem for designed mobile minimachine and detail of components in mainframe. For minimize of mechanical looses are all shafts supported by ball bearings. There are eight single line radial bearings.
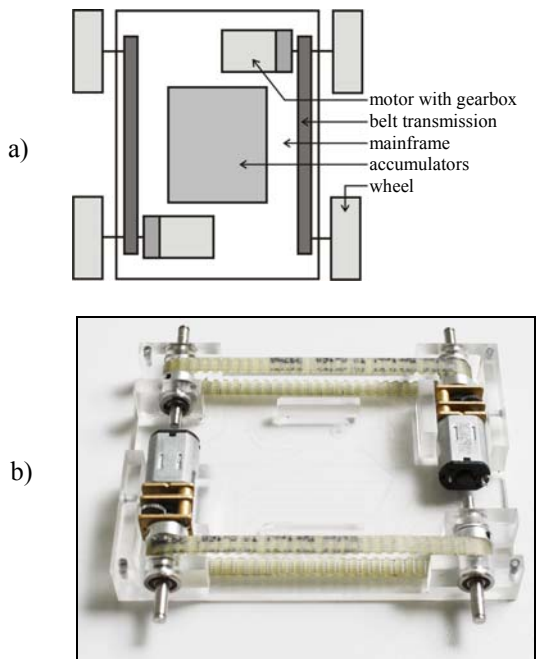
Fig. 3. Propulsive system in mainframe

a) Schematic diagram of mainframe, b) propulsion system (motors and belts transmissions) in mainframe

Main advantages of this conception are simplicity and possibility of turn round on place (zero radius of turn). Disadvantage of system without steered wheels is higher driving resistance during turns. This disadvantage is highly reduced by using of wheg components. In mainframe are also placed accumulators, shown on Fig. 3a. DC motors with four stage gearbox with gear ratio 1:100 have torque 0,15 Nm and maximal rotation speed 144 rpm. Powered shafts have also circles for incremental sensors of rotation, which allows feedback for precise positioning of mobile minimachine.

III. CONTROL ELECTRONICS CONCEPT

Control electronics of developed prototype is situated on upper part of mainframe (Fig. 4).
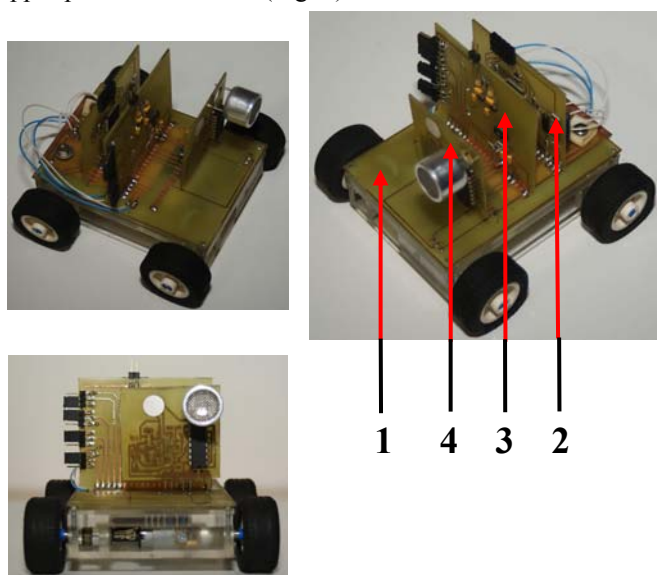
Fig. 4. Chassis of mobile minimachine with control electronics (1) interconnection part, (2) microprocessor part, (3) power supply part, (4) sensor part

It consists of interconnection part (1) with electrical interconnections and power driver for motors. On this part are connected function parts - microprocessor part (2), power supply part (3) and sensor part (4). Conception of interconnection part allows simple expansion by add-on next function parts or replaces existing parts for enhanced with extending functionality. Interconnections part and other function parts are made on printed circuit boards.

Microprocessor part, schematically shown on Fig. 5, consists of microcontroller DS89C430 with crystal and support circuits for in-system programming.
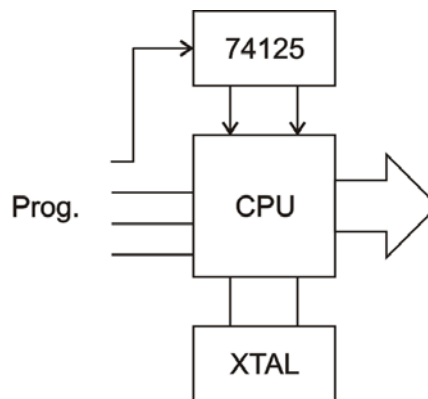
Fig. 5. Block diagram of microprocessor part

Power supply part, schematically shown on Fig. 6, consists of voltage stabilization for electronics of other function parts, circuits for charging accumulators in mainframe and switch for disconnection accumulators during charge.
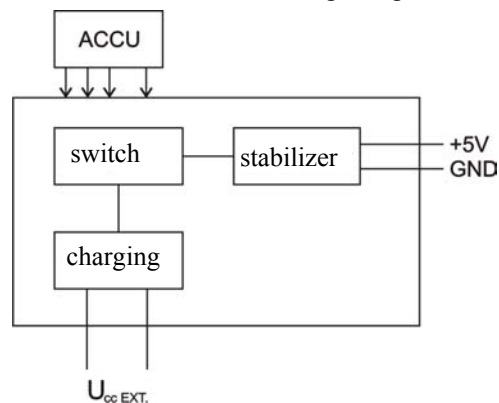
Fig. 6. Block diagram of power supply part

Sensor part, schematically shown on Fig. 7, consists of circuits for power driving of piezoelectric speaker, amplifier for signal scanned from z piezoelectric sensor and detector. Generation of signal and detector signal processing are implemented in microcontroller DS89C430 on microprocessor part.
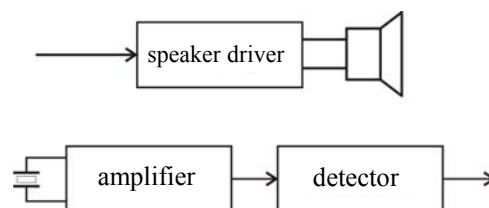
Fig. 7. Block diagram of sensor part

For development of modified solution concept of mobile minimachine is necessary to satisfy following requirements: suitable technology of integration for control electronics with possibility of heat dissipation from power components and location of sensors on proper position within the chassis of mobile minimachine. Consequently, there is requirement for integration of power supply (3) on interconnection part (1) (Fig. 8).

- interconnection part with power components
- microprocessor block
- sensor block

3D system integration
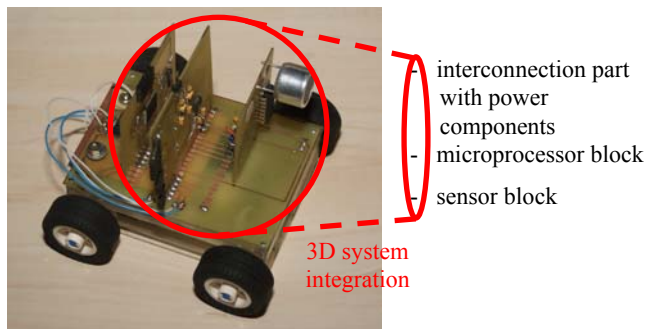
Fig.8. Integration of control electronics for mobile minimachine

Based on previous requirements was designed new concept of integration. For new integration concept was used heterostructure LTCC structure – HTCC substrate, which is based on different thermal conductivity of used materials, where HTCC substrate works as cooler [3]. Proposed design count with heterostructe LTCC structure – HTCC substrate with shaped LTCC structure. During the design process is necessary to solve technological problems with shaping fixture, design of topology for electrical circuits with focus on sensor location and possibility of local heat dissipation from power components. Simultaneously are solved problems with connection of function blocks, gripping and contacting of piezoelectric sensor.

For designed control electronics was milled precise slot (1) on top side of chassis. This slot is used for gripping of control electronics above cooling channel (2), which allows air flow around bottom side of heterostructure (Fig. 9).

(1)

(2)
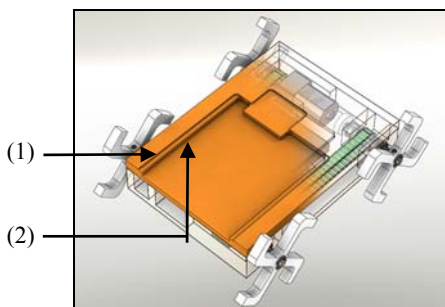
Fig.9. Slot (1) for gripping of control electronic and cooling channel (2)

IV. CONCLUSION

Article gives brief preview on processed design solution of autonomous mobile minimachine for varied terrains. Actual design state is focused on construction of alternative chassis and testing of control electronics built on ceramic substrates.

REFERENCES

[1] S, Slosarčík, I. Vehec: TECHNOLÓGIA LTCC. VIENALA 2010, ISBN 978-80-89232-97-0.
[2] S. Slosarčík, I. Vehec, P. Cabúk, M. Jurčišin: Technology and application of 3D shaped LTCC modules for pressure sensors and microsystems. In: Journal of Microelectronics and Electronic Packaging, ISSN 1551-4897, vol. 6, no. 3 (2009), p. 158-163.
[3] S. Slosarčík, P. Cabúk, S. Kardoš: Heteroštruktúra korundový substrát – 3D LTCC modul. Patentová prihláška č.PP000362010.

Model of the autonomic transport robot with optical and ultrasonic sensor based on system Lego Mindstorms NXT 2.0

Marek ŠTULAJTER

Supervisor of bachelor work: prof. Ing. Dobroslav Kováč, CSc.

Department of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

marek.stulajter@student.tuke.sk

Abstract - This paper describes an autonomic mobile transport robot with optical and ultrasonic sensor system based on Lego Mindstorms NXT 2.0. This model represents a device that would be able to carry different segments of cars in their manufacture. The transport-robot follows with optical sensor monitors the line marked in contrasting colors on the floor, for example black color on white background or reversal. The ultrasonic sensor can avoid collision with unexpected objects that may appear before him.

Keywords **- mobile transport robot, control unit, sensors, actuators, software.**

I. INTRODUCTION

Nowadays automatization is becoming an irreplaceable part not only in car producing industry. Sophisticated automatic devices are replacing manual labor of the individuals. In car production, either passenger vehicle or freight vehicle are used different devices for transport such as conveyor belts or carts with rails. This model of the robot could replace such vehicles. Robot movements will be led by the contrast line marked on the floor using an optical light sensor, which monitors the track, and ultrasonic sensor will avoid the collision with unexpected obstacles. In such way designed assembly line may be easily altered depending on the immediate needs of production. For example, when changing the vehicle model way it will produce another product. If necessary it is easy to change the leading track of transport only with repainting line on the floor. With carts running on rails is change of track impossible. Thus created workspace is very adaptable to the needs of the manufacturer. Moreover if into the track of vehicle gets an object that could cause conflict ultrasonic sensor detects it and the vehicle stop.

II. DESCRIPTION OF THE DEVICE

The device consists of an intelligent NXT brick that is actually the brain of the robot, two servo motors, optical (color) sensor that monitors the line, and ultrasonic sensor for monitoring whether there is an obstacle in front of the vehicle.

A. NXT Intelligent brick

The main part of a smart cube is 32-bit ARM7 microprocessor type AT91SAM7S256 Atmel with 48MHz operating frequency, internal 256 kb sized flash memory where are stored individual programs, and 64kb of RAM memory. The microprocessor works with 8bit coprocessor with 4 kilobytes flash memory, 512 bit ram memory and clocked at 8 MHz. The brick has four inputs numbered 1234, here are connected sensors and three outputs ABC, here are connected servomotors. Sensors and motors are connected with six line cable and RJ-12 connector. This connector is different from classical RJ-12 connector it has the lock on right side and classical connector has it in the middle.

Another part of the brick is the display and four buttons. The display is LCD with a resolution of 100 x 64 pixels and displays 8 lines, two buttons are used to orientate into the menu to the left or to the right and the other two are used to confirm the operation and make step back. Brick can be connected to a PC using a USB cable or Bluetooth. To supply smart cube serves 6 AA batteries (1.5 V), or a lithium battery with a capacity of 1400 mAh.
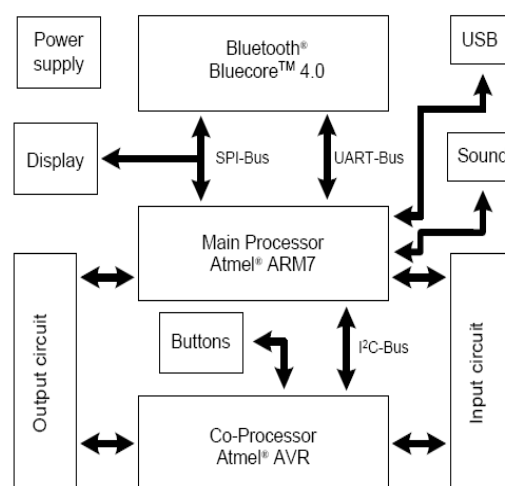
Fig. 1. Block diagram of the intelligent brick

B. Servomotors

Each servomotor has built-in rotation sensor. The servomotor can be controlled very precisely after one degree in the range from 0 ° to 360 ° what provides very precise

control. Servomotor without load reaches 170 rpm and maximum torque 50N.cm. Low speed and power of engine is the result of internal gearing. 9 Volts batteries supply the motors and in the unloaded condition at 170 rpm each servomotor takes 60 mA. In the maximum stage of loading when rotor does not rotate each engine takes 2A. Such high current servomotor can withstand only in a short time therefore it is protected with thermistor that limits the maximum current flowing into the engine.
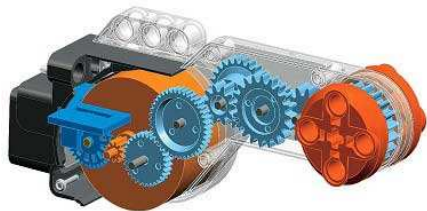
Fig. 2. Look inside the servomotor

Programming of servomotor in the NXT-G program is simple. It can be used Move or Motor icons. Using the Move icon you can adjust both motors speeds and this is used in mastering direction of vehicle only with different speeds and powers of motors. This command determines the direction of track of the vehicle if it turns left, right or goes straight. It also can be managed the strength, duration in seconds, or rotation of the engine in degrees or rotation speed. This is monitored by the built in speed sensor and for one revolution it is considered 360 degrees. The last option is the Unlimited duration of rotation till unexpected damage or some specified condition occurs. Command Motor has the same job as the Move but it manages only one servomotor.
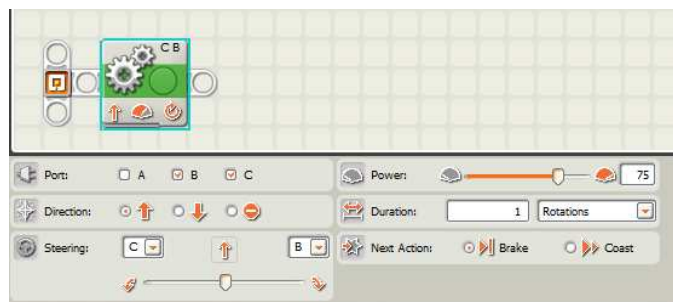
Fig. 3. Order move
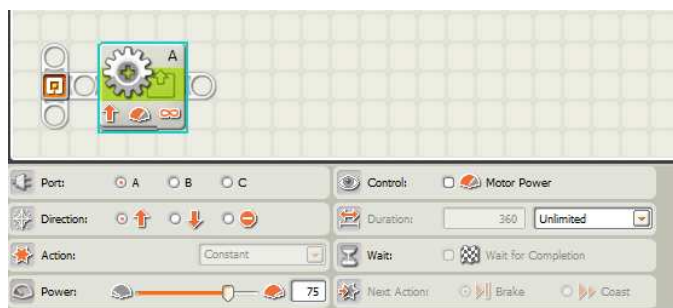
Fig. 4. Order motor

C. Optical sensor

The color sensor is used as the optical sensor. It works as a sensor that detects six colors. Can recognize light intensity in the room and surfaces or may work as a lamp emitting red, green and blue light. The sensor uses RGB LED diodes that

sequentially emit light on subject and then evaluates the reflected light. The sensor is sensitive to all wavelengths. It connects to port 3 the optimal position for sensor is in distance 1 cm above the surface and should be placed perpendicularly otherwise the data may be inaccurate.

Fig. 5. Color optical sensor

Light sensor adjusts much like a servomotor. In program for a mobile robot it is used in branching working in light sensor mode it is possible to set so. The first step of programming was to load value that has displayed when the sensor was above pure black background - so it was the black line value. The value obtained was somewhere around 25 it depended on the inequalities of the track so that was why the setting was on 30. On a white background value was about 50 depended on whether there was completely clean white surface or there were shades of some different colors. When the sensor was partly above the black line and partly above the white background it had to get to the threshold value. This value was ranged somewhere around 27 + - 2 therefore it was chosen 30 as the threshold value.
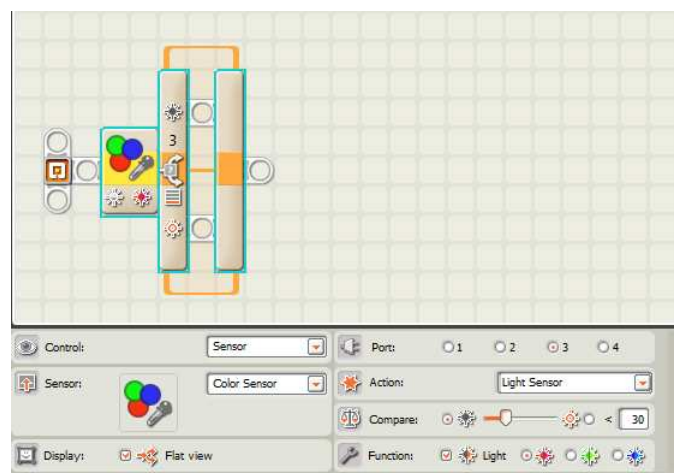
Fig. 6. Branching with a color sensor, working in fashion light sensor

D. Ultrasonic sensor

This sensor works by sending and receiving sound waves. It consists of two eyes, while the left eye serves as a signal transmitter and the right eye serves as a receiver. The sensor sends a sound wave into the space that is reflected from an object located in front of the sensor and returning to the sensor that it directly. Based on the time which has passed from posting to the adoption of signal may determine the distance in which is object located. The maximum distance to which it

can operate is 255 cm with an accuracy of +/- 3 cm. Distances less than 3 cm cannot be measured. The sensor is connected to port 4. For recognize are the best articles from the large solid materials from which the wave reflects well. Contrast poorly identify round objects made of soft materials, also looking for harder courses, which are thin or small. For best operation should be placed in a horizontal position when it is placed at angle that distorts the field.

Fig. 7. Ultrasonic sensor

Like a light sensor works in the branching, so that the sensor operates in a loop, which is performed using an ultrasonic sensor. The program can adjust the distance that it exceeds the loop will cease to be repeated. In another program may be adjusted such that it takes up another sub, it's all individual, the possibilities are many. Distance is set in cm or inches and is also set whether to consider a smaller or larger distance than is set.
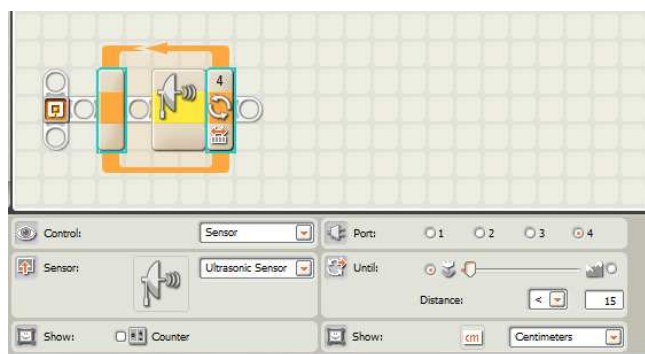
Fig. 8. Ultrasonic sensor

III. PROGRAM

On programming the NXT brick, there are several programs such as graphic NXT-G, which is part of a kit, or a variety of text editors based on the C language as BricxCC. NXT-G program is user-friendly, the programs are moving to block orders, where such movement is adjusted values engines or loop where the repetition of the conditions set by sensor, time, or counter again and again. You can track the current values of sensors and motors. Commands are performed in the order they are arranged (stacked) in a row. The disadvantage of this program environment is the breadth of programs. The increasing number of orders, the whole program becomes opaque. Good programming environment is BricxCC it is classical text editor, which use language based on C language.

The first part of the program is monitoring the area ahead of the vehicle with ultrasound sensor. At the beginning of the repetitive loop is based on data from the ultrasonic sensor. The value which is sensor taking how limit is 15 cm. This loop will be executed until the value of the distance will be greater than is set in the program, in this case it is 15cm, then implementing a program to follow the black line. When the

value falls below the 15 cm loop is completed and made to order to stop the engine. Sensor is connected to port 4.

The second part of the program is watching the black line. The program begins with branching, which is performed on the basis of color sensor working in a light sensor mode. The sensor is connected to port 3. Light value should be set somewhere in the middle of the reference range. Minimum value is the value loaded from the center line and the maximum value is the value for pure white background. Robot tries to follow the line to keep the left edge of the look from the perspective of the robot. If the value is greater than the set in this case 30 is performed rotation of the motors connected to port B with a power of 10 and subsequently rotate the motors connected to port C to the 70th power. If the value drops below 30 shall be turning the engine on the port B with a force of 70 and subsequently rotate the engine with the power of the 10. The resulting movement is not smooth but the robot follows the line nicely.
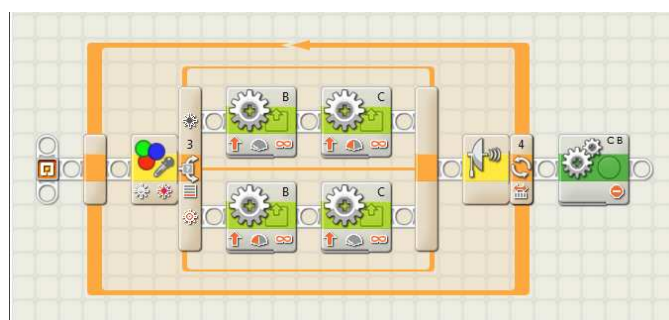
Fig. 9. Program

IV. CONCLUSION

The result of this work is a functional model of an autonomous transport robot. The system works fairly reliably, it can be gradually improved and expanded by rotating an ultrasonic sensor which could be identified objects in the space. For example, could come around the object located on the line before him.

REFERENCES

[1] NXT user guide (Electronically PDF format)
[2] Lego Technic Motors compared characteristics. Available on internet: http://www.philohome.com/motors/motorcomp.htm
[3] The mindstorms NXT review. What's inside the NXT brick. Available on internet: http://mindstormsnxt.blogspot.com/2006/08/whats-inside-nxt-brick.html

Multilevel Converter for Photovoltaic System

[1]Marek PÁSTOR, [2]Marcel BODOR, [3]Erik EÖTVÖS

[1, 2, 3]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]marek.pastor@tuke.sk, [2]marcel.bodor@tuke.sk, [3]erik.eotvos@tuke.sk

Abstract—**The main objective of this paper is to describe a cascade H-bridge inverter with focus on photovoltaic applications. The designed 15-level cascade inverter is controlled by mean of RT-Lab. The current control technique is used. The current regulator is designed with the help of Rapid Control Prototyping. Experimental results are briefly described as well.**

Keywords—**cascade H-bridge inverter, current control, photovoltaics, Rapid Control Prototyping.**

I. INTRODUCTION

The solar energy and especially photovoltaics is one of the fastest growing industries in the world. There is a demand for high quality electrical energy and thus the use of photovoltaics is almost impossible without modern power electronics. If we omit the simplest PV battery charger there always has to be certain power conditioning unit (PCU) between the PV generator and the load whether to maximize the energy yield or to change certain qualities of the electrical energy. Whether it is a stand alone PV electrical generator or a grid connected system there is a demand to change the DC voltage to the AC voltage, to maximize the energy yield and to monitor the whole system. This is done by the mean of a PV inverter. The use of the PV inverter is to change the DC voltage to the AC voltage and to adapt the PV generator to the electrical load as well as to monitor the whole system. There are several types of PV inverters according to the topology. This paper describes the cascade H-bridge inverter which can be used for photovoltaic applications. The cascade H-bridge inverter is an alternative to the single H-bridge inverter in photovoltaic systems. The cascade inverter can produce lower THD of the grid current and THD of the output voltage, requires smaller filters, can transfer more power and has smaller du/dt stresses. There is a need to increase the lifetime of photovoltaic inverters as well as their reliability. High voltage stresses decrease the lifetime of many electrical components [4]. Lower du/dt stresses of components in multilevel H-bridge inverter can help to meet these needs.

II. CASCADE H-BRIDGE INVERTER

A. Basics

Cascade inverters belong to the multilevel power converters. Multilevel power converters are mainly used for medium and high power application due to utilization of several power semiconductor switches with separated DC sources connected in series. Multilevel power converters have several advantages over single level power converters [1]:

staircase output voltage, low common mode voltage, low distortion input current, and lower switching frequency. The disadvantages are higher number of power semiconductor switches, more complex control technique and higher conduction losses.

B. Cascade H-bridge inverter

A single-phase structure of a general 7-level cascade H-bridge inverter is shown in Fig.1. The nominal power of the proposed cascade H-bridge inverter is 3600W. Maximal power depends on used IGBTs and heat sinks.
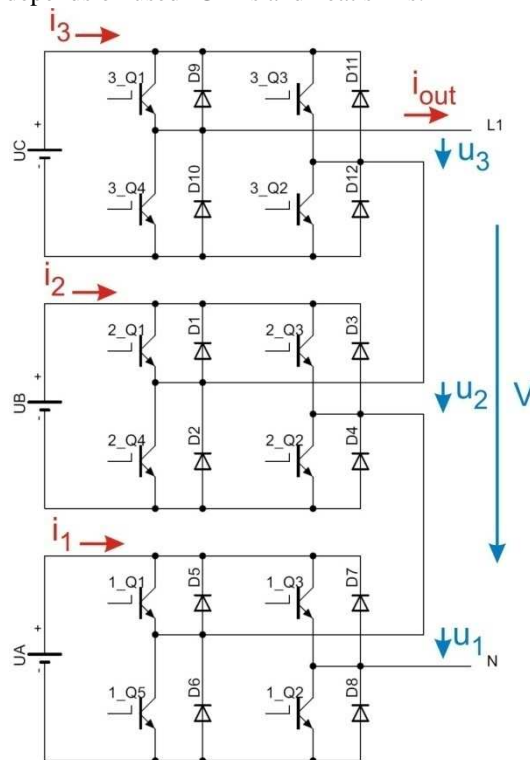
Fig. 1. Single-phase cascade H-bridge inverter with three separated DC sources (U$_A$ = 240V, U$_B$ = 120V, and U$_C$ = 60V), capable to create 15 voltage levels at its output.

The number of output phase voltage levels n is defined by:

$$n = 2d + 1 \tag{1}$$

where:
d – is the number of separated DC sources.

However it is possible to create more voltage levels at the output of the cascade inverter. Each H-bridge converter can create positive, negative or zero voltage on its output with magnitude equal to the DC source. Thus there are 15 possible combinations for the cascade H-bridge inverter with 3 separated DC sources.

C. Output voltage control technique

There are several methods for a voltage control of the cascade inverter. One of them is the sinusoidal PWM from high switching frequency PWM modulation strategies [1].

The amplitude modulation index for the multilevel inverter is defined by:

$$m_a = \frac{A_m}{(n-1)A_C} \qquad (2)$$

where:
A_m – is the modulation signal amplitude,
A_C – is the carrier signal amplitude,
n – is the output voltage level number.

The frequency modulation index is defined by:

$$m_f = \frac{f_C}{f_m} \qquad (3)$$

where:
f_C – is the frequency of the carrier signal,
f_m – is the frequency of the modulation signal.
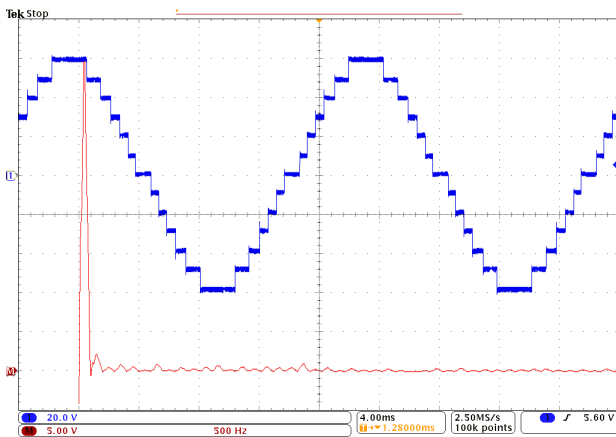
Fig. 2. Measured output voltage of 15-level (3 DC sources: U_A = 40 V, U_B = 20 V, U_C =10 V) cascade H-bridge inverter with voltage control (m_a=0.8, m_f=2, U_{RMS}=42V, THDu=9%).

The partial voltages at the output of each H-bridge inverter are shown in Fig.3. It can be clearly seen that each H-bridge inverter is switching with different frequency, which is increasing as the voltage of the H-bridge inverter is decreasing.
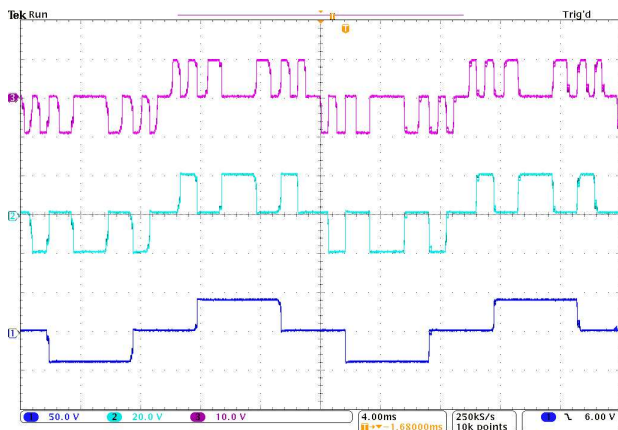
Fig. 3. Measured partial output voltages u_1, u_2, u_3 of the cascaded H-bridge inverter (amplitudes: 40, 20 and 10V), m_a = 0,8, m_f = 2

D. Current control technique

If we consider the DC/DC converter at the inverter's input this DC/DC converter acts as a voltage source. Thus the inverter must be a voltage source inverter (VSI). There are two main control strategies for VSI: the voltage control (VCVSI) and the current control (CCVSI). They vary in the way they control the power flow. The VCVSI uses the control of the decoupling inductor voltage to control the power flow and the CCVSI uses the decoupling inductor current to control the power flow. The CCVSI is faster, can control active and reactive power flow independently but can not provide the voltage support to the load, can not operate without the grid. The CCVSI can be used for power factor correction due to the fact, that it can control the reactive power independently [2]. It also has a limited short circuit current compared to the VCVSI.

There are various techniques how to archive the current control in CCVSI. One of them is a predictive current control for voltage source inverters [3].

The easiest case is to use a simple RL filter to decouple the grid voltage E and the inverter's output voltage V. For circuit in Fig.4 it can be written:

$$\overline{V} = R\overline{I} + L\frac{d\overline{I}}{dt} + \overline{E} \qquad (4)$$
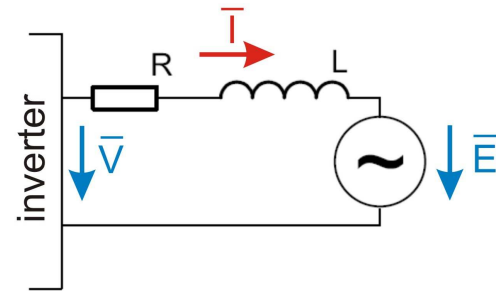
Fig. 4. The RL filter between the inverter's output and the grid used to decouple the output voltage and the grid and to filter higher harmonics.

If we consider the sampling period T to be sufficient small and the vectors $\overline{V}$ and $\overline{E}$ are constant between two sampling periods, current from (4) can be discretized as follows [3]:

$$\overline{I}_{(k+1)T} = e^{-\frac{R}{L}T}\overline{I}_{kT} + \frac{1}{R}\left(1 - e^{-\frac{R}{L}T}\right)\left(\overline{V}_{kT} - \overline{E}_{kT}\right) \qquad (5)$$

The current value $I(k+1)T$ is predicted by the Lagrange quadratic extrapolation.

III. CONTROL OF CASCADE INVERTER

The after mentioned current control technique was used to control the laboratory model of a 15-level (three DC sources: 240V, 120V and 60V) cascade inverter. The current regulator was designed using Rapid Control Prototyping technique (RCP) with help of Matlab/Simulink and RT-Lab. When using RCP, which is a part of the Hardawre in the Loop simulation, the regulator is simulated on a computer in real time and is connected to a real plant. This technique can easily verify the controller design. The plant is the 15-level cascade inverter with 3 separated DC sources in this case. Simulink

was used to create a program scheme of the regulator and RT-Lab was used to simulate the regulator in real time.
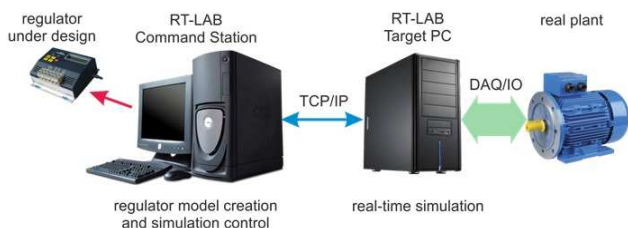
Fig.5. Interconnection of simulated regulator and real plant using RT-Lab during RCP procedure

The connection between the real-time system and a real hardware is accomplished by a mean of DAQ card with digital and analog inputs/outputs. The dead time needed to prevent short circuit in H-bridge leg is realized by the hardware driver circuit. The current regulator is sampled with sample time T. With current system configuration the sample time can be set as low as 50 µs.

A. *Experimental results of the CCVSI*

The current control technique was tested with the laboratory model of a 15-level (three DC sources: 240V, 120V and 60V) cascade inverter. The cascade inverter was not connected to the grid. Thus the E_{kT} in (5) was equal to zero. The cascade inverter was loaded with RL load (R = 13.2Ω, L = 6mH). The sample time of the current controller was set to T = 200 µs. The THDi of the load current and the efficiency of the cascade inverter were measured.

When evaluating behaviour of the cascade inverter with the suggested current control it is more accurate to consider sampling frequency of the current controller rather than the switching frequency of the inverter. The controller chooses the best voltage vector at the inverter's output and thus each H-bridge switches with different frequency.
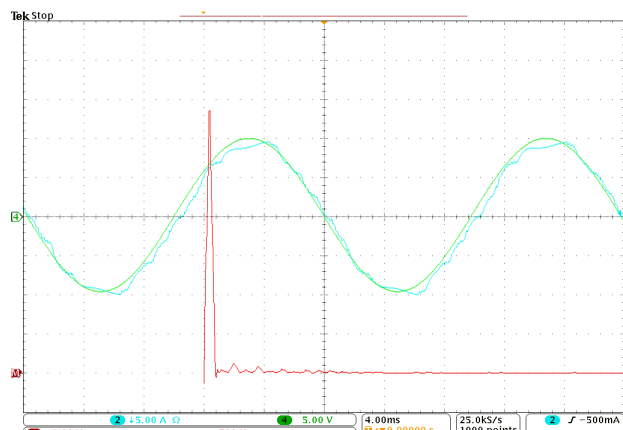
Fig. 6. Output current of cascade inverter with current control: CH1: required value (Imax = 10A), CH4: real value, M: FFT up to 1kHz (T = 200µs, k1 = 0,5, k2 = 0,027, R = 13,2Ω, L = 6mH, U_A = 240V, U_B = 120V, U_C = 60V

In Fig.6. the real output current of the cascade inverter together with desired value inverter are shown. Inaccuracy of current tracking is caused by fluctuations in DC supply voltages of separate DC sources at the H-bridge inverters' inputs. This fact is documented in Fig. 7. where the output voltage of the cascade inverter is shown.

The THDi of the output current at the output of the cascade inverter when supplying the RL load was measured. The results are shown in Fig. 8. If the DC supply voltage was better filtered the THDi would be lower.
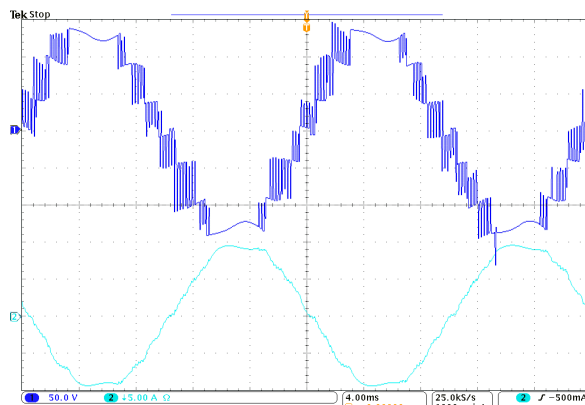
Fig. 7. Output current and voltage of cascade inverter with current control: CH1: output voltage, CH2: output current (Imax = 10A),T = 200µs, k1 = 0,5, k2 = 0,027, R = 13,2Ω, L = 6mH, U_A = 240V, U_B = 120V, U_C = 60V
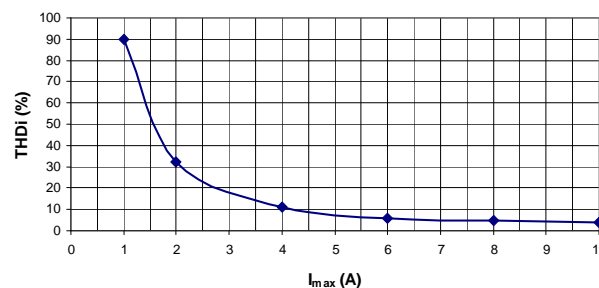
Fig. 8. THDi of the output current of the cascade inverterT = 200µs, k1 = 0,5, k2 = 0,027, R = 13,2Ω, L = 6mH, U_A = 240V, U_B = 120V, U_C = 60V

IV. CONCLUSION

Control by mean of RT-Lab is not possible in real cascade inverter. The DSC control system is under development now as well as the connection of the inverter to the grid. Using filter with higher order is also one of the aims of future work.

REFERENCES

[1] S. Khomfoi, L. M. Tolbert, "Multilevel Power Converters" from M. H.Rashid, "Power Electronics Handbook", Second edition, Elsevier, 2007,ISBN10: 0-12-088479-8, pp. 451–482.

[2] S. H. Ko, S. R. Lee, H. Dehbonei, Ch. Nayar, "A Comparative Study of the Voltage Controlled and Current Controlled Voltage Source Inverter for the Distributed Generation System" [Online]. Available: http://www.itee.uq.edu.au/~aupec/aupec05/AUPEC2005/Volume1/S073.pdf

[3] G. S. Perantzakis, F. H. Xepaps,S. A. Papathanassiou, S. N. Manias, "A Predictive Current Control Technique for Three-Level NPC Voltage Source Inverter", [Online].

[4] "Current Demand of High Performance Inverters for Renewable Energy Systems" [Online]. Available: http://www.gpec.dee.ufc.br/publicacoes/Sdaher_Antunes.pdf

[5] J. R. Espinoza, "Inverters" from M. H.Rashid, "Power Electronics Handbook", Second edition, Elsevier, 2007,ISBN10: 0-12-088479-8, pp. 353–404.

[6] L. M. Tolbert, F. Z. Peng, "Multilevel Converters as a Utility Interface for Renewable Energy Sources" [Online].

[7] Hamar, J., Jardan, K.,R., Nagy, I., Ohsaki, H.: Virtual Laboratory for Combined Solar Energy System. EPE 2007, Aalborg, Denmark, 2-5 September, ISBN: 9789075815108, pp. p.1-p.10 – CD.

14

Nanoscale electric circuits and the Brownian motion

¹*Gabriela VASZIOVÁ,* ²*Lukáš GLOD*

¹Dept. of Physics, FEI TU of Košice, Slovak Republic
²Dept. of Mathematics and Physics, Institute of Humanitarian and Technological Sciences,
The University of Security Management, Košice, Slovak Republic

¹gabriela.vasziova@tuke.sk, ²lukas.glod@vsbm.sk

Abstract— **Mesoscopic and nanoscale electric circuits are intensively studied due to the rapidly increasing number of technological applications [1]. Exploring the mathematical analogy between the Brownian motion (BM) and electric circuits, which are at nanoscales described by the generalized Langevin equation (LE), we calculate the fluctuations of charge and current in *RLC* circuits that are in contact with the thermal bath. Both in the movement of a colloidal particle in a solvent or in the case of nanoscale circuits the memory effects should be taken into account. As a result, the ordinary LE for the particle position or for the temporal evolution of the charge in a circuit becomes inapplicable [2]. It changes to a Volterra-type integro-differential equation with the stochastic force represented by a colored noise, as distinct from the white-noise force in the standard LE [1]. Our approach to the solution of such generalized LE is exact in the classical approximation and equally applicable to electric circuits and the BM with memory.**

Keywords— **nanoscale electric circuits, thermal fluctuations, Brownian motion.**

I. INTRODUCTION

The Brownian motion is traditionally regarded as discovered by Scottish botanist Robert Brown [3] in 1827. While Brown was studying pollen particles floating in water in the microscope, he observed minute particles in the pollen grains executing the jittery motion. After repeating the experiment with particles of dust, he concluded that the motion was due to pollen being "alive" but the origin of the motion remained unexplained. In fact similar observations had been recorded in France, in 1827, by Adolphe Brongniart [4]. However, some scientists believe that the first description of the BM [5] belongs to the Dutch physiologist Jan Ingen-Housz, best known for having discovered photosynthesis. But most probably the Brownian motion was observed soon after the discovery of microscope (the early 1590s). A remarkable description of BM of dust particles in the air is found in the philosophical poem De rerum natura by Titus Lucretius Carus (c. 60 BC), who uses it also as a proof of the existence of atoms [6]. In a series of comprehensive experiments [7], in the late 1880s, Gouy demonstrated convincingly that the irregular motion of suspended particles was not a result of external vibration, temperature, incident light, surface tension, etc., in

other words that BM was indeed a fundamental physical property of fluid matter. The first person to give a theory of BM was Louis Bachelier in 1900 in his PhD thesis "The theory of speculation". Early attempts to quantify anything in Brownian motion had met with great difficulties. In fact, before the parallel theoretical treatments of Einstein, Smoluchowski, and Langevin, the real problem was that the experimentalists did not know quite what to measure. But after 1905, theory, even when incompletely comprehended, provided a framework that guided the experimentalists toward the most meaningful measurements. Einstein obtained [8] the now-familiar expression showing that the mean square displacement of a particle from some origin increases with the square root of time; and into this expression he was able to insert his relation for the diffusion coefficient. Meanwhile Marian von Smoluchowski [9] obtained essentially the same expression for the time dependence of displacement as Einstein, though with a numerically different coefficient (later found to be in error by Langevin). The third derivation of the time dependence of the diffusion coefficient was made by Langevin [10] and was, according to his own words, "infinitely more simple" theory of BM than the Einstein's one. It was Langevin's friend, Jean Perrin, who finally provided the determining measurements. In this contribution we show, how the generalized form of the LE, that takes into account the possible memory effects in the evolution of the system, can be easily solved.

II. THE BROWNIAN MOTION WITH MEMORY

In [11], we have explored the correspondence between the motion of the Brownian particles (BP) and the fluctuations in the electric circuits, which are in contact with the thermal bath. When the BP moves in a liquid as in the experiments [12], for the correct description of the particle behavior (especially at short times) the inertial effects during the motion must be taken into account. This means that not only the particle mass should be nonzero, but also the memory effects play a role [13] (the state of the particle motion at the time t depends on the particle velocities and accelerations in the preceding moments of time). Mathematically it displays in the generalization of the Langevin equation, which becomes an

integro-differential equation [2]

$$M\dot{\upsilon} + \int_0^t \Gamma(t-t')\upsilon(t')dt' = \eta(t), \qquad (1)$$

where, for the BM, M could be the mass of the BP with the velocity $\upsilon(t) = \dot{x}(t)$. The memory in the system is described by the kernel $\Gamma(t)$, and $\eta(t)$ is a stochastic force with zero mean. The fluctuation-dissipation relation dictates that the condition $\langle \eta(0)\eta(t) \rangle = k_B T \Gamma(t)$ at $t > 0$ must be satisfied [14]. We are interested in finding the time-dependent diffusion coefficient $D(t) = V(t)/2$, while the mean square displacement (MSD) of the particle is given by $\xi(t) = \int_0^t V(\tau)d\tau$. The initial conditions are $\xi(0) = \dot{\xi}(0) = 0$, also the condition $\dot{V}(0) = 2k_B T / M$ must hold. It follows from the equation that concretizes the Vladimirsky rule [15] for this special case,

$$M\dot{V} + \int_0^t \Gamma(t-t')V(t')dt' = 2k_B T. \qquad (2)$$

Now let us consider an example, in which the memory kernel is specified. We chose $\Gamma(t)$ in the Ornstein-Uhlenbeck form $\Gamma(t) = (\gamma^2/m)\exp(-\gamma t/m)$. Then the force $\eta(t) = m\dot{u}(t)$ corresponds to the solution of the usual LE $m\dot{u}(t) + \gamma u(t) = f(t)$. Here $u(t)$ can be the velocity of a BP, γ is a constant friction coefficient, and $f(t)$ the stochastic (white noise) force. The motion of the BPs with the mass M can be thus interpreted as being induced by the force η caused by the particles of mass m.

The Laplace transformation of Eq. (2) yields

$$\tilde{V}(s) = \frac{2k_B T}{Ms}\left(s + \frac{\gamma}{m}\right)\left(s^2 + \frac{\gamma}{m}s + \frac{\gamma^2}{mM}\right)^{-1}. \qquad (3)$$

The denominator in this expression can be given the form $(s - \alpha_1)(s - \alpha_2)$, $\alpha_{1,2} = (\gamma/2m)\left(1 \mp \sqrt{1 - 4m/M}\right)$, so that the solution is expressed as

$$\tilde{V}(s) = \frac{2k_B T}{M}\frac{1}{\alpha_2 - \alpha_1}\left(1 + \frac{\gamma}{m}\frac{1}{s}\right)\left(\frac{1}{s - \alpha_2} - \frac{1}{s - \alpha_1}\right). \qquad (4)$$

Using the tables of Laplace transforms [16] we get

$$V(t) = \frac{2k_B T}{M}\left\{\frac{\gamma}{m\alpha_1\alpha_2} + \frac{1}{\alpha_2 - \alpha_1} \times \right.$$
$$\left. \times \left[\left(1 + \frac{\gamma}{m\alpha_2}\right)\exp(\alpha_2 t) - \left(1 + \frac{\gamma}{m\alpha_1}\right)\exp(\alpha_1 t)\right]\right\}. \qquad (5)$$

The MSD is obtained by simple integration of this equation. It can be used that $\alpha_1\alpha_2 = \gamma^2/mM$; it is then seen that the first term in {} is the Einstein long-time limit $2k_B T/\gamma$.

If we denote $\mu = \sqrt{1 - 4m/M}$, $D(t)$ can be given a compact form

$$D(t) = \frac{k_B T}{\gamma}\left\{1 - \frac{1}{4\mu}\exp(-\gamma t/2m) \times \right.$$
$$\left. \times \left[(1 + \mu)^2 \exp\left(\frac{\gamma\mu t}{2m}\right) - (1 - \mu)^2 \exp\left(-\frac{\gamma\mu t}{2m}\right)\right]\right\}. \qquad (6)$$

If we forget that the motion of the particles was interpreted as induced by particles with smaller mass m, an interesting result follows from Eq. (5) in the case when $M < 4m$, i.e. when the roots α are complex. Then the solution describes damped oscillations; e.g. for $D(t)$ we have

$$D(t) = \frac{k_B T}{\gamma}\left\{1 - \exp\left(-\frac{\gamma t}{2m}\right) \times \left[\cos\left(\frac{\gamma t}{2m}\sqrt{\frac{4m}{M} - 1}\right) - \right.\right.$$
$$\left.\left. - \frac{2m/M - 1}{\sqrt{4m/M - 1}}\sin\left(\frac{\gamma t}{2m}\sqrt{\frac{4m}{M} - 1}\right)\right]\right\}. \qquad (7)$$

III. THERMAL FLUCTUATIONS IN NANOSCALE ELECTRIC CIRCUITS

The presented theory can be equally applied to a different kind of problems, the fluctuations in electric circuits. It is well known that the fluctuations in electric circuits that are in contact with the thermal bath and the Brownian motion of particles can be described by essentially the same mathematics, the standard LE of motion [17], that is appropriate asymptotically for long times, but gives incorrect dependencies of the relevant time correlation functions at short times. Thus it is natural to expect that the effects of memory exist also in electric circuits. The theory for nanoscale RLC circuits in contact with the thermal bath has been developed recently [1]. The bath was modeled by harmonic oscillators (linear LC circuits) attached to the studied circuit. After the standard canonical quantization scheme applied to the whole closed Hamiltonian system and next tracing out the bath, since only the degrees of freedom of the initial circuit are considered to be observable, the final generalized LE for the charge $Q(t)$ can be obtained. Let the resistor and the capacitor have the resistance R and the capacitance C, Γ is the maximal frequency of the bath and $\eta(t)$ is the noise, which, in the classical limit considered here, has the property $\langle \eta(t)\eta(0) \rangle = k_B T R \Gamma \exp(-\Gamma t)$. Then the equation for $Q(t)$ is [1]

$$L\ddot{Q} + \left(R\Gamma + \frac{1}{C}\right)Q - R\Gamma^2 \int_0^t d\tau \exp[-\Gamma(t-\tau)]Q(\tau) = \eta(t). \qquad (8)$$

This equation can be solved for the quantity $v(t) = d\xi(t)/dt = (d/dt)\langle[Q(t) - Q(0)]^2\rangle$. Using the rule due to Vladimirsky [15], it is just necessary to replace in Eq. (8) $Q(t)$ with $\xi(t) = \int_0^t v(t')dt'$, and the force $\eta(t)$ with $2k_B T$. The new integro-differential Volterra-type equation is now deterministic and can be solved using the Laplace transformation. For the Laplace transformed quantity $\tilde{v}(s) = \Lambda\{v(t)\}$, taking into

account the conditions $\xi(0) = v(0) = 0$ and applying the convolution theorem, we obtain the following equation:

$$\tilde{v}(s) = 2k_B T \frac{\Gamma + s}{Ls^3 + \Gamma Ls^2 + s(\Gamma R + 1/C) + \Gamma/C} =$$
$$= 2k_B T \frac{\Gamma + s}{(s - s_1)(s - s_2)(s - s_3)}. \tag{9}$$

The solution of Eq. (9) is straightforward. The inverse Laplace transform can be obtained from this expression after its decomposition to simple fractions $\sim 1/(s - s_i)$, where s_i are the roots of the cubic equation $Ls^3 + \Gamma Ls^2 + s(\Gamma R + 1/C) + \Gamma/C = 0$. Then, e.g., if the roots s_i are different, $v(t)$ will be a sum of exponentials, $v(t) = \sum_{i=1}^{3} A_i \exp(s_i t)$. It is seen from the decomposition

$$\frac{\Gamma + s}{(s - s_1)(s - s_2)(s - s_3)} = \frac{A_1}{s - s_1} + \frac{A_2}{s - s_2} + \frac{A_3}{s - s_3}, \tag{10}$$

where the constant $A_1 = (\Gamma + s)/[(s_1 - s_2)(s_1 - s_3)]$ is obtained after multiplying this equation by $(s - s_1)$ and then setting $s = s_1$; other constants are of the same form with the cyclic change of the indexes $1\rightarrow2\rightarrow3\rightarrow1$. The limits of long and short times are determined by the limiting behavior of $\tilde{v}(s)$ at $s \rightarrow 0$ and $s \rightarrow \infty$, respectively. In this way we find $v(t) \approx (2k_B T/\Gamma L)[1 - \exp(-\Gamma t)] \approx 2k_B Tt/L$, as $t \rightarrow 0$. This corresponds to $\xi(t) \sim t^2$, exactly as in the case of MSD for the BPs. The autocorrelation function for the current, $i(t) = \langle I(t)I(0)\rangle$, exponentially approaches the constant $k_B T/L$, as $t \rightarrow 0$. At long times $v(t)$ and $i(t)$ converge to zero with the relaxation time $RC + 1/\Gamma$,

$$v(t) \approx 2k_B TC(RC + 1/\Gamma)^{-1} \exp[-t/(RC + 1/\Gamma)].$$

The "MSD" $\xi(t)$ at long times approaches the value $2k_B TC$. If $L = 0$ from the beginning, the condition $\xi(0) = 0$ cannot be satisfied, since the equation of motion would yield $0 = 2k_B T$. If $L \neq 0$, at $t = 0$ we have $L\ddot{\xi}(0) = 2k_B T$, so that the expansion at small t is $\xi(t) = k_B Tt^2/L + ...$

IV. CONCLUSION

The applicability of the Langevin equation to the Brownian motion in suspensions has a strong limitation. When the characteristic times in experiments are comparable to or smaller than $R^2\rho/\eta$, where R is the particle radius and ρ and η are the density and viscosity of the solvent, the viscous aftereffect should be taken into account. For such time scales the solution of standard LE becomes inapplicable. A similar situation takes place for nanoscale electric circuits, which are now possible to realize experimentally. In both cases the evolution of the studied systems is characterized by some kind of memory. The corresponding Langevin equations thus turn to stochastic integro-differential equations and involved calculations are needed to solve them. Our work represents a

progress with such problems. It consists in converting the complicated stochastic equations of motion into the deterministic ones that are quite easily solvable analytically. We have shown that within the classical consideration the exact solution of the generalized Langevin equation can be found. The method is directly applicable to the description of the Brownian motion driven by a colored noise since for the usual micron-sized particles the classical approach is relevant. However, for nanoscale circuits and low temperatures the quantum effects become important [1]. The future work thus requires a generalization of the presented method of solution to the case of quantum noise.

REFERENCES

[1] A. E. Allahverdyan, Th. M. Nieuwenhuizen, "On testing the violation of the Clausius inequality in nanoscale electric circuits", *Phys. Rev. B* 66, 115309 (2002) (see also comment on this article in: E. P. Gyftopoulos, M. R. von Spakovsky, Comments on testing the violation of the Clausius inequality in nanoscale electric circuits, *arXiv* 0706.2842v1).

[2] W. T. Coffey, Yu. P. Kalmykov, J. T. Waldron, *The Langevin Equation. With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*, World Scientific, New Jersey e.a., 2nd Ed. (2005).

[3] R. Brown, "A brief account of microscopical observations made in the months of June, July, and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies", privately printed (1827), and reprinted in *Edinb. New Philos. J.* 5, 358 (1828).

[4] A. K. Brongniart, "Memoire sur la génération et le développment de l'embryon dans les végétaux phanérogames", *Ann. Sci. Naturelles* 12, 41 (1827).

[5] J. Ingen-Housz, "Remarks on the use of the microscope" (in German), in *Vermischte Schriften phisisch-medizinischen Inhalts [von] J. Ingen-Housz*, 2nd ed., vol. 2, Niklas Karl Molitor, Transl. and Ed. Vienna: C. F. Wappler, 122-126 (1784).

[6] Lucretius, *On the Nature of the Universe*, R. E. Latham, transl. London: Penguin Books (1994).

[7] L. Gouy, "Note sur le mouvement brownien", *J. Physique Serie II* 7, 561 (1888).

[8] A. Einstein, "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen", *Annalen der Physik* 17, 549 (1905).

[9] M. Smoluchowski, "Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen", *Ann. Phys.* 21, 772 (1906).

[10] P. Langevin, "Sur la theorie du mouvement brownien", *C. R. Acad. Sci.*, Paris 146, 530 (1908).

[11] G. Vasziova, V. Lisy, "Fluctuations in electric circuits and the Brownian motion of particles", *SCYR 2011: proceedings from conference*, Košice: FEI TU, 153-156 (2010).

[12] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, "Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales", *Phys. Rev. Lett.* 89, 050601 (2002).

[13] V. Lisy, J. Tothova, "On the (hydrodynamic) memory in the theory of Brownian motion". *arXiv:cond-mat*/0410222 (12 pp.).

[14] R. M. Mazo, *Brownian Motion. Fluctuations, Dynamics, and Applications*, New York: Oxford University Press (2009).

[15] V. V. Vladimirsky, "To the question of the evaluation of mean products of two quantities, related to different moments of time, in statistical mechanics". *Zhur. Ekper. Teor. Fiz.* 12, 199-202 (1942) (in Russian).

[16] M. Abramowitz, I. A. Stegun, *Handbook of Mathematical Functions*, Washington, DC: National Bureau of Standards (1964).

[17] R. van Zon, S. Ciliberto, E. G. D. Cohen, "Power and heat fluctuation theorems for electric circuit", *Phys. Rev. Lett.*, 92, 130601 (2004).

Numerical Estimations of Instantaneous Power Components Electric Circuit with the Diode

[1]*Mariia MALIAKOVA,* [2]*Andrey KALINOV,* [3]*Radoslav BUČKO*

[1]Dept. of Automation Control Systems and Electric Drive, Kremenchuk Mykhaylo Ostrograskyi National University, Ukraine

[2]Dept. of Automation Control Systems and Electric Drive, Kremenchuk Mykhaylo Ostrograskyi National University, Ukraine

[3] Dept. of Theoretical Electrical Engineering and Electrical Measurement, Technical University of Košice, Slovakia

[1]marry_88@mail.ru, [2]scentr@kdu.edu.ua, [3]radoslav.bucko@tuke.sk

Abstract—**In this paper the instantaneous power method which based on the forming and calculation systems of instantaneous power balances equations was used for numerical estimation of instantaneous power components on the diode. The instantaneous power components balances or the electric circuit researched was forming with the automated method of instantaneous power components getting. Also the analysis of getting results is presents.**

Keywords—**Instantaneous power, automated algorithm, electric circuit, analysis, diode.**

I. INTRODUCTION

Today the semiconductor elements are integral components of electrical devices and systems. The relative methods of mathematical analysis are required for researching the physical process of these systems, especially for the high-power systems).

In the paper [1] the instantaneous power (IP) technique (IPT) is offered for analyzing the systems with semiconductor elements. The paper [4] shows that IP components should be taken into account when forming the IP balances, even on the ideal diode [3]. During the voltage and current spectral analysis on the ideal diode, the voltage and current harmonic components (including a constant component) should be obtained. This causes the presence of the corresponding IP components during the symbolic analysis of the equations. Although physically IP value on the ideal diode are equal to zero over a whole period during the voltage and current instantaneous values multiplication.

The approximation of the nonlinearity is always attended by certain errors. Thus it is reasonable to make a research of the approximation accuracy influence on the forming of IP components balances.

Numerical estimation of IP components of the electric circuit with the diode applying the IP components balances is the objectives of this paper.

II. ELECTRIC CIRCUIT RESEARCHED

Let's analyze the IP components balances of the electric circuit with the diode connecting in series with an active resistance (fig. 1). An automated method of IP components forming is used for the IP balances constructing [2].
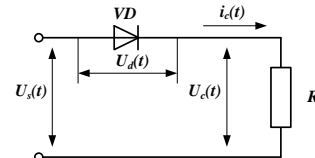
Fig. 1. Electric circuit researched

A polynomial function is selected as an approximating function of the diode nonlinear characteristic. It connects the output Uc and input Us voltage on the diode. Several different order polynomials $Uc(Us)= a_0+a_1Us +a_2Us^2+...+a_nUs^n$ (fig. 2) are analyzed for getting the maximum accurate dependence.

There are obtained coefficients of represented polynomials: 3-rd order $a_0=1.054\cdot10^{-1}$, $a_1=5.757\cdot10^{-1}$, $a_2=4.252\cdot10^{-1}$,

a)

b)

c)
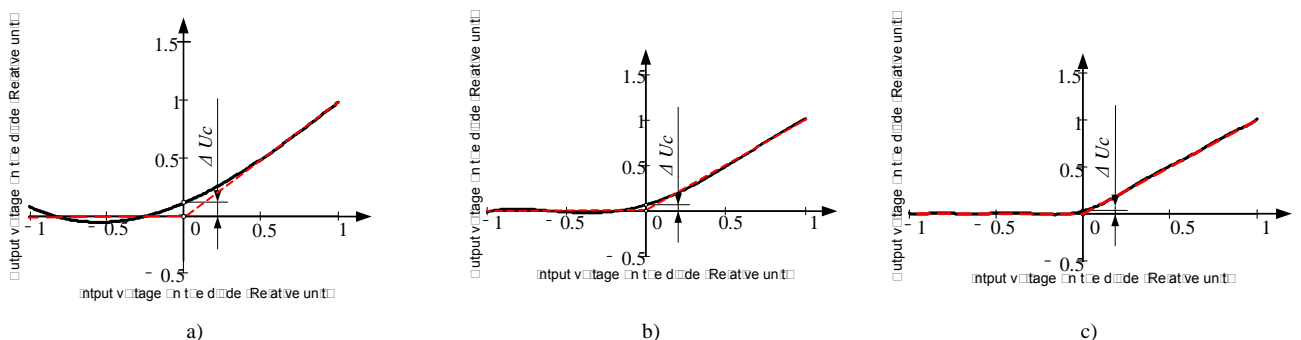
Fig. 2. Approximation dependence of the nonlinear characteristic of the diode, presented by the polynomial of a) 3-rd, b) 6-th and c) 10-th order, (---) – nonlinear characteristic of the ideal diode, (—) – approximation dependence $a_3=-9.727\cdot10^{-2}$, $a_4=-3.525\cdot10^{-1}$, $a_5=9.662\cdot10^{-2}$, $a_6=2.791\cdot10^{-2}$ and 10-th order $a_0=2.441\cdot10^{-2}$, $a_1=5.000\cdot10^{-1}$, $a_2=1.886$, $a_3=-6.925\cdot10^{-12}$, $a_4=-6.536$, $a_5=2.562\cdot10^{-11}$, $a_6=1.356\cdot10^{1}$, $a_7=-3.482\cdot10^{-11}$, $a_8=-1.322\cdot10^{1}$, $a_9=1.564\cdot10^{-11}$, $a_{10}=4.789$.

Calculated parameters of researched electric circuit (fig.1) were compared with parameters getting on the mathematical model, created in the environment of Matlab. There the active resistance value $R=1$ Om and input voltage value $Us_{a1}=1$ V were used. Using the next dependence $Ud=Us_{a1}-Uc$ the reverse voltage on the diode was calculated. Curve of changing Ud
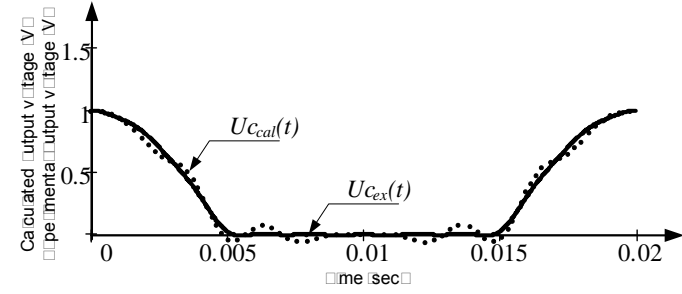
shown on figure 3.

III. ACCURACY EVALUATION OF THE GETTING RESULTS

As criteria for evaluating the exactness of calculation electric circuit with the diode were adopted: a relative error by the effective value or power and current signals, and comparing the degree of matching model and calculated curves (fig. 4) for the coefficient of determination R^2 and the standard deviation (SD).

The results obtained from the comparison given in Table I.
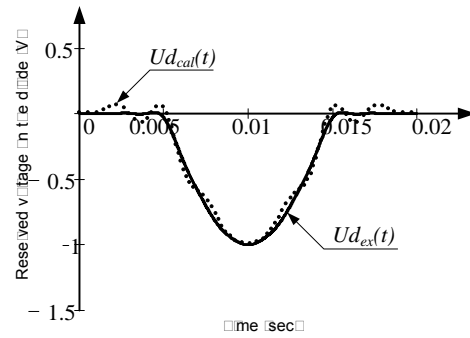
a)   b)

Fig. 3. Curve of a) output voltage on the diode $Uc_{cal}(t)$ – curve, constructed by the calculated values of output voltage on the diode, $Uc_{ex}(t)$ – curve, constructed by the values of output voltage on the diode, getting with the experimental way; b) voltage on the diode $Ud_{cal}(t)$ – curve, constructed by the calculated values of reserved voltage on the diode, $Ud_{cal}(t)$ – curve, constructed by the values of reserved voltage on the diode, getting with the experimental way
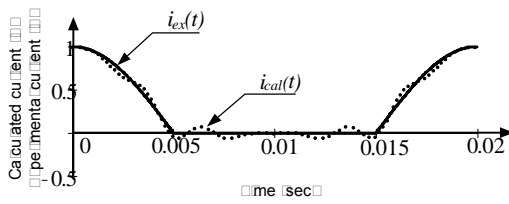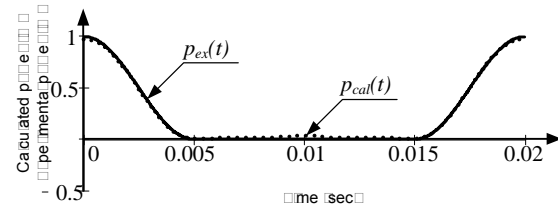
a)   b)

Fig. 4. Curve recovery of a) $i(t)$, b) $p(t)$ with the approximation of the diode nonlinear dependence with 10-th order polynomial, here $i_{cal}(t)$ – curve, constructed by the calculated values of current; $i_{ex}(t)$ – curve, constructed by the values of current, getting with the experimental way; $p_{cal}(t)$ – curve, constructed by the calculated values of power; $p_{ex}(t)$ – curve, constructed by the values of power, getting with the experimental way

TABLE I
EXACTNESS EVALUATION OF THE CALCULATION OF CURRENT AND IP COMPONENTS, WHERE THE NONLINEAR CHARACTERISTIC OF THE DIODE PRESENTED AS 3-TH, 6-TH AND 10-TH ORDER POLYNOMIAL

3-rd order polynomial			
Δi_{ef}, %		Δp_{ef}, %	
15.575		16.635	
SD	R^2	SD	R^2
55.068	0.815	49.036	0.803
6-th order polynomial			
Δi_{ef}, %		Δp_{ef}, %	
15.226		2.304	
SD	R^2	SD	R^2
29.307	0.901	0.915	0.996
10-th order polynomial			
Δi_{ef}, %		Δp_{ef}, %	
1.678		2.079	
SD	R^2	SD	R^2
3.461	0.988	0.426	0.998

IV. NUMERICAL ESTIMATION OF IP COMPONENTS ELECTRIC CIRCUIT WITH THE DIODE

It is necessary to mark, that calculated power harmonics onthe diode are changing with the change of the accuracy of the approximation dependence of the diode nonlinear characteristic. To study the IP harmonic components on the diode at different orders of approximation function (3-th and 6-th and 10-th order) they are correlated with the IP value at the power source, which is adopted as basic. During research the corresponding constant power components and root-mean-square values with and without constant components are analyzed (table 2).

TABLE II
NUMERICAL ESTIMATION OF IP COMPONENTS ON THE DIODE

Order of approximation polynomial function	$\dfrac{P_{0d}}{P_{0s}}$	$\dfrac{\sqrt{\sum\limits_{k=0}^{10} P_{kd}^2}}{\sqrt{\sum\limits_{k=0}^{10} P_{ks}^2}}$	$\dfrac{\sqrt{\sum\limits_{k=1}^{10} P_{kd}^2}}{\sqrt{\sum\limits_{k=1}^{10} P_{ks}^2}}$
3-rd order	0.571	0.443	0.38
6-th order	0.077	0.066	0.05
10-th order	0.046	0.039	0.03

P_{0d}, P_{0s} – IP constant component on the diode, power supply, accordingly; $\sqrt{\sum\limits_{k=0}^{10} P_{kd}^2}$, $\sqrt{\sum\limits_{k=0}^{10} P_{ks}^2}$ – quadratic IP value on the diode, power supply, accordingly, with constant component; $\sqrt{\sum\limits_{k=1}^{10} P_{kd}^2}$, $\sqrt{\sum\limits_{k=1}^{10} P_{ks}^2}$ – quadratic IP value on the diode, power supply, accordingly, without constant component; k – IP harmonic number.

There is the IP components balances of electric circuit with the diode in numerical form, where approximation of the nonlinear characteristic of the diode, presented by the polynomial of 10-th order.

Table 2 shows, that the constant power component on the diode is the most influential. This fact can be explained as follows: the biggest errors by approximation occurs in the small-values region of the input voltage (fig. 2). This errors reduced to the value of output voltage will be equal to $\varDelta Uc$, even when the input voltage $Uc(t)$ equal to zero. This value is going down with the increase in the approximation function orders. The same dynamic is observed in the constant IP components (table 2).

It is logically to make the conclusion, that the numerical IP harmonic components on the diode is as less as higher the order of approximation function of the nonlinear characteristic of the diode.

$$p_{Ra}(t) + p_{da}(t) = p_{sa}(t) =$$

$$
\begin{pmatrix}
0.477 \\
0.405 \\
0.240 \\
0.831e\text{-}1 \\
-0.100e\text{-}2 \\
-0.174e\text{-}1 \\
0.102e\text{-}1 \\
0.141e\text{-}1 \\
-0.809e\text{-}2 \\
-0.118e\text{-}1 \\
-0.288e\text{-}2
\end{pmatrix}
+
\begin{pmatrix}
0.229e\text{-}1 \\
-0.275e\text{-}10 \\
0.955e\text{-}2 \\
-0.182e\text{-}10 \\
0.101e\text{-}2 \\
-0.673e\text{-}11 \\
-0.102e\text{-}1 \\
-0.141e\text{-}1 \\
0.809e\text{-}2 \\
0.118e\text{-}1 \\
0.288e\text{-}2
\end{pmatrix}
=
\begin{pmatrix}
0.5 \\
0.405 \\
0.25 \\
0.831e\text{-}1 \\
0.129e\text{-}10 \\
-0.174e\text{-}1 \\
0.317e\text{-}11 \\
0 \\
0 \\
0 \\
0
\end{pmatrix}.
$$

14

The identity of represented balances of IP components proves the accuracy and efficiency of the analysis of the electric circuit with diode applying the automated method of IP components forming [2].

V. CONCLUSIONS

After the obtained data have been analyzed, it is reasonably to make the next conclusions: first, presence of IP components on the diode is required for the identity of IP components balances; second, the value of the researched IP components on the diode depends on the quality accuracy of the approximation of the diode nonlinear characteristic.

ACKNOWLEDGEMENT

REFERENCES

[1] Калінов А.П., Малякова М.С. Розрахунок електричного кола з діодом методом миттєвої потужності / Вісник КДУ імені Михайла Остроградського.– Кременчук: КДУ імені Михайла Остроградського. – 2010. – Вип. 4(63). – Ч. 3. – С. 31-35.

[2] Калінов А.П., Малякова М.С. Автоматизований алгоритм розрахунку електричних кіл за складовими миттєвої потужності / Електромеханічні і енергозберігаючі системи. – Кременчук: КДПУ імені Михайла Остроградського, ІЕЕКТ. – 2009. – Вип. 1. – С. 34-38.

[3] Родькин Д.И. Баланс составляющих мгновенной мощности полигармонических сигналов / Вісник КДПУ. – Кременчук: Кременчуцький державний політехнічний університет. – 2007. – Вип. 3(44). – Ч. 1. – С. 66 – 77.

[4] Родькин Д.И., Калинов А.П. Энергопроцессы в цепях с несинусоидальными напряжениями и токами без парадоксов (комментарий к дискуссионной статье профессора Долбни В.Т. «Об одном парадоксе, возникающем при анализе цепей с выпрямителями») / Вісник НТУ «ХПІ», «Проблеми автоматизованого електропривода. Теорія і практика». – Харків: НТУ «ХПІ». – 2010. – №28. – С. 590-599.

On the colored noise generated by a nonlinear Langevin equation

[1]*Lukáš GLOD,* [2]*Gabriela VASZIOVÁ*

[1]Dept. of Mathematics and Physics, Institute of Humanitarian and Technological Sciences,
The University of Security Management, Košice, Slovak Republic
[2]Dept. of Physics, FEI TU of Košice, Slovak Republic

[1]lukas.glod@vsbm.sk, [2]gabriela.vasziova@tuke.sk

Abstract—**The Langevin equation, originally designed to describe the Brownian motion of particles in suspensions, can be considered also as a generator of colored noise. In the present contribution one example of such generator is studied. The dissipative friction force is taken from the current models of Brownian motors and is nonlinear. In this case the spectrum of the generated noise is not known but some its properties can be found from a simple dimensional analysis. Namely, the relation between the intensities of the resulting noise and the driving white noise is obtained when the friction factor in the dissipative force depends on the velocity of the particle as ~ $\upsilon^{2\alpha}$.**

Keywords—**Langevin equation, diffusion, noise generator, nonlinear friction.**

I. INTRODUCTION

In the article published in Comptes Rendus in 1908 Paul Langevin proposed a different approach to description of Brownian motion (BM) than that of Einstein and Smoluchowski. It was assumed to be "infinitely simpler" than the Einstein´s one and seemed to be based only on the equipartition theorem. As distinct from Einstein who derived and solved a partial differential equation for the probability density of a Brownian particle (BP) [1], Langevin used Newton's second law and incorporated in it a random force driving the particle. The obtained ordinary differential equation gave rise to the theory of stochastic processes and is now widely used to describe Markov (memoryless) processes in different fields of physics, chemistry, electrical engineering, biology, and even in finance and social sciences [2].

Let us consider the motion of the BP in a fluid. The Langevin equation describing the BM of a particle with the mass m is

$$m\dot{\upsilon} = -\gamma\upsilon + \sqrt{2D}\xi(t), \qquad (1)$$

where $\upsilon = dr/dt$ is the particle velocity, the irregular impulses from the surrounding molecules are described by the (white noise) force ~ $\xi(t)$ with zero mean, which has the statistical properties $\langle\xi_i(t)\xi_j(t')\rangle = \delta_{ij}\delta(t-t')$ and the intensity $D = k_B T\gamma$ (k_B is the Boltzmann constant, T is the temperature and the friction coefficient γ for a spherical particle with the radius R is the Stokes one, $\gamma = 6\pi R\eta$, η is

the dynamic viscosity of the surrounding fluid), and the diffusion coefficient of the BP represents the simplest realization of the fluctuation-dissipation theorem (FDT), $D_{eff} = D\gamma^{-2} = k_B T/\gamma$ (this formula, usually called the Stokes-Einstein relation, should carry the name of Sutherland as well [3]). Equation (1) is easily solved with these properties. The derivation of the mean square displacement (MSD) from Eq. (1) is very simple [4] and can be found in many textbooks and monographs. The MSD for any of the coordinates x, y, and z is defined as $X(t) = \langle[x(t) - x(0)]^2\rangle$, where $\langle...\rangle$ denotes the statistical averaging. As distinct from the Einstein relation $X(t) = 2D_{eff}t$, which is valid only for "infinite" times t, Eq. (1) predicts the behavior of the particle for arbitrary time intervals:

$$X(t) = 2D_{eff}t\left(1 - \frac{1 - \exp(-\gamma t/m)}{\gamma t/m}\right). \qquad (2)$$

However, this solution has (as concerns BM) a very limited applicability. In fact, it is true only for heavy particles in a low-density environment (such as dust particles in a gas) at short times, or alternatively at long times, when, however, the Einstein theory is valid. In the following section the case when the friction coefficient γ is replaced by the function of velocity $\gamma(\upsilon) \approx \upsilon^{2\alpha}$ will be considered. Although the solution for the spectrum has not yet been obtained, some interesting conclusions about the noise intensity can be done based on a simple dimensional analysis.

II. LANGEVIN EQUATION WITH NONLINEAR FRICTION

The standard BM is described by the Langevin equation (1), in which the dissipative Stokes´ force is linearly-dependent on the velocity of the BP. This kind of friction has a very limited use. In a more general case, the friction force against the movement of the BP depends nonlinearly on the velocity, and can also be a function of the position of the particle. A number of stochastic processes in physics, electrotechnics, biology and other sciences is described by analogous equations.

Various generalizations of Eq. (1) during the past 20 years

were used to describe the motion of living objects – molecular motors, crawling cells, flocks of animals, that are definitely out of thermodynamic equilibrium.

Consider now the case when the coefficient of friction is a function of velocity [5]

$$m\dot{v} = -\gamma(v)v + \sqrt{2D}\xi(t).$$ (3)

Since the Brownian particle is in equilibrium with the liquid the noise intensity is related to temperature and friction coefficient by $D = k_B T\gamma$. A typical choice for the nonlinear friction function is $\gamma(v) = v^2 - v_0^2$ (Rayleigh-Helmholtz friction), which is interesting because of the fact that at $|v| < v_0$ the friction function acts as an energy pump. Another generalization, studied e.g. in [5], is for a d-dimensional ($d = 1, 2, 3$) system with the friction function

$$\gamma(v) = \gamma v^{2\alpha},$$ (4)

where γ and α ($\alpha = 0, 1, 2, 3$) are constants. The Langevin equation thus becomes

$$m\dot{v} = -\gamma v^{2\alpha} v + \sqrt{2D}\xi(t).$$ (5)

Eq. (5) has no solution in elementary functions [6]. If we derive the Fokker-Planck equation corresponding to Eq. (5), we can calculate the stationary distribution function for the Brownian motion process described by the Langevin equation (5).

For the nonlinear Langevin equation (5) the drift coefficient $D^{(1)}$ and the diffusion coefficient $D^{(2)}$ are given by

$$D^{(1)}(v,t) = -\frac{\gamma}{m}v^{2\alpha+1}, \quad D^{(2)}(v,t) = \frac{D}{m^2}.$$ (6)

The corresponding Fokker-Planck equation thus has the form

$$\frac{\partial}{\partial t}P(v,t) = \frac{\gamma}{m}\frac{\partial}{\partial v}\left[v^{2\alpha+1}P(v,t)\right] + \frac{D}{m^2}\frac{\partial^2}{\partial v^2}P(v,t),$$ (7)

which has the stationary solution in the form

$$P_{st}(v) = N\exp\left[-\frac{m\gamma v^{2\alpha+2}}{2D(\alpha+1)}\right],$$ (8)

where N is the normalization constant. From the normalization condition $\int_{v_{min}}^{v_{max}} dv P_{st}(v) = 1$ and for the natural boundary conditions ($v_{min} = -\infty$, $v_{max} = +\infty$) we therefore have

$$P_{st}(v) = \frac{1}{2}\frac{(2\alpha+2)^{\frac{2\alpha+1}{2\alpha+2}}}{\Gamma(1/(2\alpha+2))}\left(\frac{m\gamma}{D}\right)^{\frac{1}{2\alpha+2}}\exp\left[-\frac{m\gamma v^{2\alpha+2}}{2D(\alpha+1)}\right].$$ (9)

III. THE DIFFUSION COEFFICIENT IN THE CASE OF NONLINEAR FRICTION

In the mentioned work [5], the "diffusion coefficient" of the particle, D_{eff}, is given the main attention. D_{eff} can be calculated from the MSD of the particle,

$$D_{eff} = \lim_{t\to\infty}\frac{1}{2td}\sum_{j=1}^{d}\left\langle\left[x_j(t) - x_j(0)\right]^2\right\rangle,$$ (10)

or using the Kubo relation for the velocity autocorrelation function (VAF),

$$D_{eff} = \frac{1}{d}\int_0^\infty d\tau\langle v(t)v(t+\tau)\rangle.$$ (11)

Knowing D_{eff} is important not only for the description of various kinds of the nonlinear BM, e.g. in physics and biology [7 - 9], but also for designing colored-noise generators [5]. Indeed, while the random force $\xi(t)$ describes the white noise with equal amplitudes of any frequency component of the force spectrum, the dynamics of Eq. (5) can be regarded as a noise generator with very different properties. In such a generator, the "diffusion coefficient" D_{eff} corresponds to the noise intensity of the "velocity" v. It is of particular interest to know the properties of this noise and its dependence on the system parameters D, γ, and m.

In the one-dimensional case, one can also calculate the factor of proportionality by means of an exact quadrature result recently derived [10]. We will demonstrate the scaling relations for the cases $\alpha = 0$ (normal BM) and $\alpha = 1$ (the high-speed limit of the Rayleigh friction), as well as for two examples of even stronger nonlinearity with $\alpha = 2$ and $\alpha = 3$. We will explain by intuitive arguments the remarkable findings that the diffusion coefficient for $\alpha = 1$ does not depend on the noise intensity at all and that it decreases with increasing noise for $\alpha > 1$.

The exact formula for the diffusion coefficient in the one-dimensional case ($d = 1$) from [10] (for the specific case $g(v) = \sqrt{D}$) reads

$$D_{eff} = \frac{\int_0^\infty dv_2 e^{U(v_2)}\left[\int_{v_2}^\infty dv_1 v_1 e^{-U(v_1)}\right]^2}{(D/m^2)\int_0^\infty dv_3 e^{-U(v_3)}},$$ (12)

where $U(v) = \gamma m v^{2\alpha+2}/2D(\alpha+1)$. Inserting the potential, introducing new variables $u_i = v_i(\gamma m/D)^{1/(2\alpha+2)}$ and carrying out the integral in the denominator, one arrives at [5]

$$D_{eff} = \frac{D^{\frac{1-\alpha}{1+\alpha}}m^{\frac{2\alpha}{1+\alpha}}}{\gamma^{\frac{2}{1+\alpha}}}\frac{(2\alpha+2)^{\frac{2\alpha+1}{2\alpha+2}}}{\Gamma(1/(2\alpha+2))}\times$$
$$\times\int_0^\infty du_2\exp\left[\frac{u_2^{2\alpha+2}}{2\alpha+2}\right]\left\{\int_{u_2}^\infty du_1\exp\left[-\frac{u_1^{2\alpha+2}}{2\alpha+2}\right]u_1\right\}^2.$$ (13)

The first factor is the desired scaling law in D and γ while the remaining form a numerical factor that still depends on the exponent α.

For $\alpha = 0$ (normal BM) one recovers the simple relations

$$D_{eff} = \gamma^{-2}D, \quad \alpha = 0.$$ (14)

For $\alpha = 1$ (the friction force is proportional to the cube of the velocity) one has

$$D_{eff} = \frac{m}{2\gamma}\Gamma\left(\frac{3}{4}\right)\int_0^\infty dx\,\text{erfc}^2(x^2)e^{x^4} \approx 0.4875 m/\gamma, \quad \alpha = 1.$$ (15)

For higher α, the double integral in (13) can be reduced to a single integral over a product of an exponential and an incomplete Gamma function. Here we just state the numerical factors resulting from the scaling law and the numerical evaluation of the remaining integral [5]

$$D_{eff} \approx 0.3737 m^{4/3}\gamma^{-2/3}D^{-1/3}, \quad \alpha = 2$$ (16)
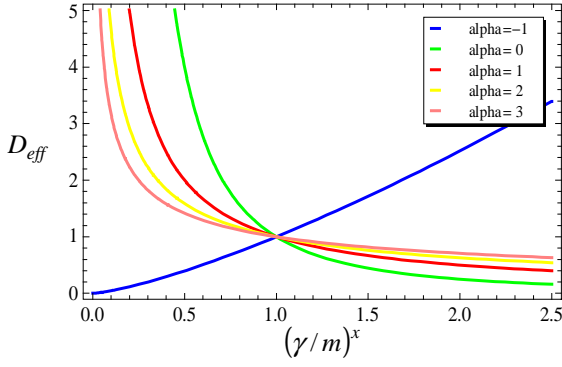
Fig. 1. Diffusion coefficient D_{eff} vs. $(\gamma/m)^x$.

Fig. 2. Diffusion coefficient D_{eff} vs. $(D/m^2)^y$.

$$D_{eff} \approx 0.3201 m^{3/2} \gamma^{-1/2} D^{-1/2}, \quad \alpha = 3. \tag{17}$$

The purpose of the work [5] (see Eq. (13)) was to show that

$$D_{eff} \approx D^\beta \gamma^{\beta'} m^{\beta''}, \tag{18}$$

where

$$\beta = \frac{1-\alpha}{1+\alpha}, \quad \beta' = -\frac{2}{1+\alpha}, \quad \beta'' = \frac{2\alpha}{1+\alpha}. \tag{19}$$

Here we use that the Langevin equation (5) is in fact governed not by three but only by two parameters, e.g. γ/m and D/m^2. If M, L, and T stay for mass, length, and time, respectively, then, using that the dimension of $\gamma v^{2\alpha}$ is M/T and D_{eff} has the dimension L^2/T, $(D) = (M^2 L^2 / T^3)$ and $(v) = (L/T)$, one easily finds from the expression

$$D_{eff} \approx \left(\frac{\gamma}{m}\right)^x \left(\frac{D}{m^2}\right)^y \tag{20}$$

the following relations for the coefficients x and y:

$$x = -\frac{2}{1+\alpha}, \quad y = \frac{1-\alpha}{1+\alpha} \quad (\alpha \neq -1) \tag{21}$$

$$x = 4/3, \quad y = -1 \quad (\alpha = -1) \tag{22}$$

so that the dependence of D_{eff} on the model parameters exactly corresponds to Eqs. (18) and (19) with the coefficients β found for $\alpha = 0, 1, 2$, and 3 [5]. Here, the dependence of D_{eff} on α is determined for arbitrary α immediately from the simple dimensional consideration. In particular, the interesting case (mentioned but not touched in [5]) when α is negative is also solved. While in [5] the question has been opened whether here the system shows normal diffusion at all, now it is seen that the answer is not. For $\alpha > 1$ and $\alpha \leq -1$ one observes that the diffusion coefficient of the BP decreases with the increase of the Langevin noise. "The exciting and unexplored problem" of the power-law friction is covered by the solution (21) as well. The relation (20) is shown in Fig. 1 and Fig. 2.

To finish with this type of the nonlinear Langevin equation note that it is appropriate to study this equation in the dimensionless form. It has been also done in [5] in numerical simulations: the graphs for dimensionless D_{eff} as a function of D have been calculated for $\alpha = 1, 2, 3$, and $d = m = \gamma = 1$. To work with dimensionless variables, one has to define the characteristic units for the mass (μ), distance (ρ), and time (τ). Then, measuring the mass, distance and time in these units, the form of Eq. (5) remains the same but all the quantities become dimensionless. Let us choose these units as those characteristic for a typical micrometer-sized BP at room temperatures in water, assuming the validity of the Stokes´ friction force. That is, $\mu = (4\pi/3) \cdot 10^{-15} kg$ and $\rho = 10^{-6} m$. Taking $\tau = m/\gamma$ to be the relaxation time for the BP, this characteristic unit will be $\tau = (2/9) \cdot 10^{-6} s$. This time corresponds to Lindner´s choice $m = \gamma = 1$ [5]. As to the noise intensity and diffusion coefficient expressed in the units for a typical BP, $\mu^2 \rho^2 / \tau^3$ and ρ^2 / τ, respectively, we have the dimensionless $D = D_{eff} = (1.38/9\pi) \cdot 10^{-6} D^*$, where $D^* = T/(300K)$ and $T(K)$ is the temperature. Thus, $D^* = 1$ yields the values for the typical BP at $300K$ (recall that the dimensional $D_{eff} = k_B T / \gamma$ and $D = k_B T \gamma$). The numerical calculations for different friction forces are in progress. Now we only note that the simulations [5] are carried out in the region of D (from 10^{-2} to 10) that is exceedingly far from that for a real BP (for such a particle the temperatures in [5] are from about $6 \cdot 10^7$ to $6 \cdot 10^{10} K$). Due to this the behavior of the simulated BP cannot be compared at all with the behavior of the typical particle.

IV. CONCLUSION

In this contribution the famous Langevin equation is considered as a generator of the colored noise. If this equation is used to describe the Brownian motion of particles, the noise is represented by the particle velocity v and its properties, in particular the color, i.e. the dependence on the frequency, are determined by the velocity autocorrelation function of the particle. We have described one example in which the Brownian particle produces a colored (correlated) noise. Instead of the Stokes friction, a force proportional to $v^{2\alpha} v$ is used; this choice is inspired by some actual models of Brownian motors [5]. The resulting Langevin equation is nonlinear and hardly solvable analytically. It is however possible to obtain some properties of the intensity of the

generated noise using just a dimensional analysis. In particular, it can be shown that for the cubic dependence of the force on υ the noise intensity does not depend at all on the intensity of the white noise driving the particle. In some cases the intensity of the noise (the particle diffusion coefficient) decreases with the increase of the white noise intensity. Since the nonlinear Langevin equation is difficult to solve analytically, currently we make attempts to solve it numerically for different friction forces. From the view on this equation as a colored noise generator, even a more important problem seems to be that of designing a generator with given desired properties of the noise.

REFERENCES

[1]   A. Einstein, "On the movement of small particles suspended in a stationary liquid demanded by the molecular kinetic theory of heat", Ann. Phys. 17, 549 (1905).

[2]   W. T. Coffey, Yu. P. Kalmykov, J. T. Waldron, *The Langevin Equation. With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*. Singapore e.a.: World Scientific, 2004.

[3]   W. Sutherland, "A Dynamical Theory for Non-Electrolytes and the Molecular Mass of Albumin", Phil. Mag. 9, 781 (1905).

[4]   P. Langevin, "On the Theory of Brownian Motion", C. R. Acad. Sci. 146, 530 (1908).

[5]   B. Lindner, "Diffusion Coefficient of a Brownian Particle with a Friction Function Given by a Power Law", J. Stat. Phys. 130, 523 (2008).

[6]   I. Kaplansky, *An Introduction to Differential Algebra*. Paris: Hermann, 1957.

[7]   P. Reimann, "Brownian Motors: Noisy Transport far from Equilibrium", Phys. Rep. 361, 57 (2002).

[8]   W. Ebeling, L. Schimansky-Geier, "Swarm dynamics, attractors and bifurcations of active Brownian motion", Eur. Phys. J. Special Topics 157, 17 (2008).

[9]   B. Lindner, E. M. Nicola, "Diffusion in different models of active Brownian motion", Eur. Phys. J. Special Topics 157, 43 (2008).

[10] B. Lindner, "The diffusion coefficient of nonlinear Brownian motion", New. J. Phys. 9, 136 (2007).

14

Phase Evolution of Solder Alloy

Juraj ĎURIŠIN

Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic

juraj.durisin@tuke.sk

Abstract—**In-situ X-ray diffraction experiments using synchrotron radiation were employed to analyze microstructure evolution of 95.5Sn3.8Ag0.7Cu (wt. %) lead-free solder alloy during heating (30-250ºC), isothermal annealing (240ºC) and cooling (250-30ºC). The special emphasis was placed at the study of melting and solidification process, explaining formation, distribution and the order of crystallization of the crystal phases (β-Sn, intermetallic compounds) in the solder alloy.**

Keywords—**X-ray diffraction, microstructure, lead-free solders, crystal phase.**

I. INTRODUCTION

Lead-free solders based on SnAgCu alloys (with melting temperature under 250°C) are extensively used as low temperature solder materials in the electrotechnical industry. Therefore solder joints play important role from view of quality and reliability of electronic devices. Significant factor defining properties of solder joints is their microstructure. Microstructure of solder joints of final electronic devices is generally influenced by applied solder alloy, process of soldering, composition of soldered materials and utility of the final device. All four factors have significant impact on phase transformations in solder joints consequently influencing their properties. Microstructure of lead-free solders consists of large β-Sn dendritic crystals with typical length of 10^1-10^2 µm. Another very important component of the solders are intermetallic compounds. Solders based on Sn with additions like Ag and Cu contain dominantly fine Ag_3Sn precipitates ($\approx$1-10^1 µm) or Cu_6Sn_5 precipitates ($\approx$1 µm). Distribution of intermetallic compounds in solder alloys is based on weight ratio of chemical elements in the solders [1].

Realized in-situ measurement of 95.5Sn3.8Ag0.7Cu [14] solder alloy illustrate behavior of the alloy during temperature increase (30-250ºC) and during temperature decrease (250-30ºC), thus simulating the soldering process. The special emphasis was placed at the observation of the phase evolution mainly during the solidification process.

II. MEASUREMENT

As a source of X-ray radiation was used for the diffraction measurements B2 bending magnet beamline of the DORIS III positron storage ring at HASYLAB/DESY (Hamburg, Germany). The beamline allows performing X-ray diffraction measurements in Debye-Scherrer geometry (Fig. 1). For our measurements the wavelength was set to λ = 0.5384 Å. The use of synchrotron radiation has numerous advantages in comparison with conventional laboratory X-ray lamps since it offers high brilliance photon beams covering relatively wide energy spectrum. To speed up the measuring, diffracted X-rays were detected on position-sensitive image plate detector OBI, covering the angle range $2\theta_{max}$ =110º.
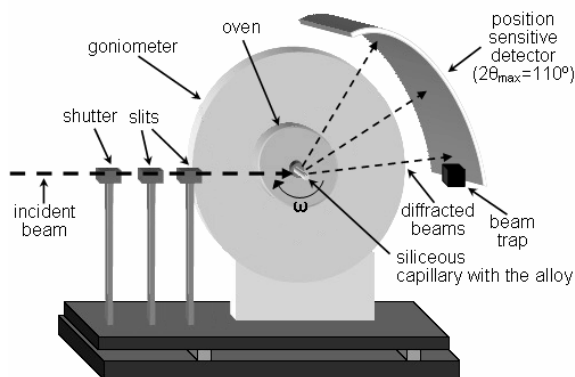
Fig. 1. Configuration of diffractometer in Debye-Scherrer geometry.

The sample of the alloy in the form of powder was placed into glass capillary having the inner diameter of 0.28 mm. The capillary containing sample was consequently mounted into the STOE furnace, thus simulating the soldering process (temperature increase and decrease). The temperature in the sample vicinity was measured by thermocouple. To eliminate the texture effect and to ensure correct intensities of the Bragg peaks, the capillary was rotating with frequency of 2 Hz (ω) around its horizontal axis, perpendicular to the incoming monochromatic beam (Fig. 1). The sample was illuminated with the monochromatic beam having the width and height of 12 and 5 mm, respectively. Measurement at every temperature took between 12-14 minutes. The temperature step of 10ºC was used during heating (30-210ºC) and cooling parts (190-30ºC), whereas the 2ºC step was applied for more detailed inspection of the melting and the solidification processes occurring in the temperature ranges (210-250ºC) and (250-190ºC), respectively.

III. RESULTS AND DISCUSSION

X-ray diffraction pattern of the fresh sample taken at 30ºC revealed the presence of the sharp and intense Bragg peaks stemming from β-Sn phase (PDF Nr. 4-673). Furthermore slight traces of Ag_3Sn crystal phase (PDF Nr. 44-1300) are clearly visible and also slight traces of Cu_6Sn_5 crystal phase

are visible (PDF Nr. 45-1488) [3]. Because our measurements were mainly aimed at evaluation the phase changes, next diffraction patterns depict behavior of the analyzed solder alloy only during the melting process (temperature increase) and at selected temperatures during the temperature decrease, where crystallization of crystal phases takes its course.

a) Heating

b) Cooling
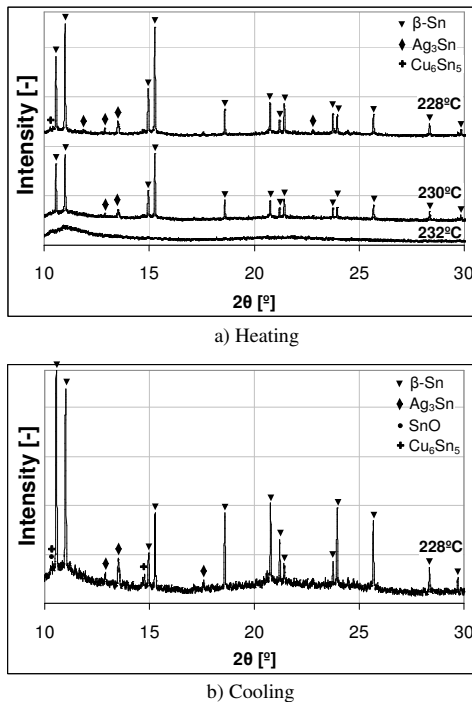
Fig. 2.   Diffraction patterns of the 95.5Sn3.8Ag0.7Cu alloy at selected temperatures.

Heating the alloy from 30ºC up to melting temperature of the alloy reveals no phase changes (phases formations). After reaching the temperature of 228ºC diffraction peaks belonging to major constituent phase β-Sn are rapidly diminishing, what is a sign of reaching the melting point of the solder. After increasing the temperature of 2°C it was possible to observe next depression of diffraction peaks intensity together with significant relative rising of the background. X-ray diffraction patterns obtained at 232°C exhibit no Bragg peaks at all and thus confirm completely molten state of the alloy (Fig. 2a).

It should be noted here that according to values published in database [2] and in binary diagrams [4], the liquidus temperature of the alloy was measured within the tolerance of +(8-12)°C. This difference is probably consequence of temperature gradient inside the furnace combined with the temperature step (2°C).

Further isothermal annealing at 250ºC for two hours revealed no phase changes and diffuse character of the X-ray diffraction (XRD) patterns remained unaffected.

In the next part we focused on the solidification process, or more precisely on crystallization of the solder alloy. XRD patterns reveal diffuse character specific for amorphous materials till temperature of 228ºC - Fig. 2b, when the first Bragg peaks corresponding to the major β-Sn phase (or also to intermetallic compounds) appear. Similar diffraction pattern (if only β-Sn reflections are taken into consideration) was

recorded during temperature increase (melting process) at the temperature of 230ºC. One may conclude here that the temperature difference between the melting liquidus and the solidification liquidus of the alloy indicates kind of thermal hysteresis.

Next table (Tab. 1) summarizes data gained from the measurement, or more precisely from XRD patterns: liquidus temperatures (for temperature increase and decrease) and also temperatures of formation of crystal phases during cooling.

TABLE I
TEMPERATURES OF LIQUIDUS AND CRYSTALLIZATION
OF CRYSTAL PHASES

	Temperature increase	Temperature decrease			
	liquidus	liquidus	β-Sn	Ag$_3$Sn	Cu$_6$Sn$_5$
95.5Sn3.8Ag0.7Cu	232ºC	228ºC	228ºC	228ºC	228ºC

IV. CONCLUSION

Diffraction measurements allowed detail analysis of the melting and solidification process of the solder alloy. Significant result of the measurements is definition of liquidus of the solder (for temperature increase and decrease) and also definition of crystallization order of crystal phases in the solder alloy, which explains distribution of the phases in lead-free solders.

REFERENCES

[1]   B. A. Cook, I. E. Anderson, J. L. Harringa, S. K. Kang: Isothermal Aging of Near-Eutectic Sn-Ag-Cu Solder Alloys and Its Effect on Electrical Resistivity, Journal of Electronic Materials, Vol. 32, No. 12, 2003.
[2]   Cooperation on Science and Technology – European Lead-free Soldering Network: Database for Properties of Lead-free Solder Alloys, February 2006.
[3]   The International Centre for Diffraction Data®: PDF-2, 1997.
[4]   ASM International's Binary Alloy Phase Diagrams, ASM International, Materials Park.

Photovoltaics as a potential primary source for meeting object´s energy demands

[1]*Martin Bačko*

[1]Dept. of Theoretical Electrical Engineering and Electrical Measurement , FEI TU of Košice, Slovak Republic

[1]martin.backo@tuke.sk

Abstract—**This article deals with photovoltaics energy as a possible primary source of meeting energy demand of an object. In first part the photovoltaics and renewable energy sources and their importance is briefly presented. Second and third part deals with photovoltaics measurements, optimal array placement. Fifth part shows the photovoltaics prices and final fifth part describes the methods for object´s energy demand assessment.**

Keywords—**Photovoltaics, Excel, Energy, Measurement**

I. INTRODUCTION

In present days, due to growing concerns about possibility of global warming, rising fossil fuel prices and the need of improved security is leading more countries to give higher priorities for integrated policies for addressing climate change and energy. The EU set an example to the rest of the world by cutting its greenhouse gas emissions by at least 20% by 2020 and final energy demand should consist from at least 20% from renewable energy sources. In term renewable sources of energy we understand ecologically clean direct or indirect form of solar energy, which can be transformed by suitable technological solution to electrical energy.. Renewable energy sources are available in much higher amount than the whole world needs and their usage is imperative for keeping the ecological equilibrium. Following figure (Fig.1) shows the solar potential of Slovak Republic as well as yearly sum of global irradiation which can be utilized by photovoltaic modules
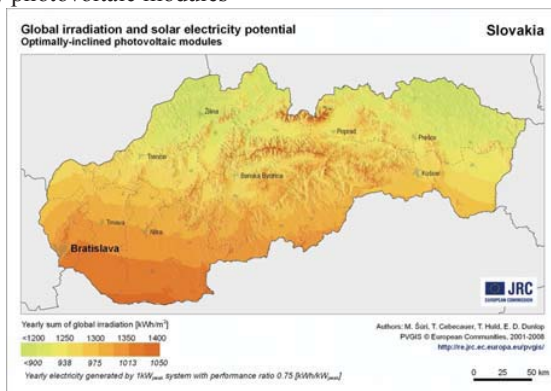
Fig.1 Global Irradiation and Solar Electricity Potential

Following figure (Fig.2) shows the photovoltaic solar electricity potential in European countries. The data was gathered by European Commision Joint Research centre.
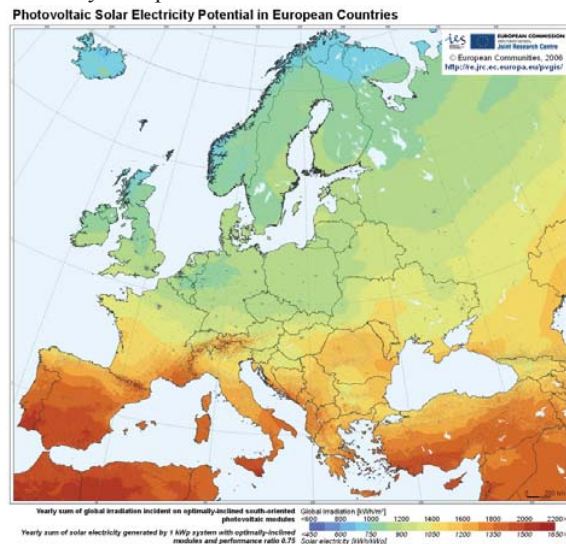
Fig.2 Photovoltaic Solar Electricity Potential in European Countries

II. PHOTOVOLTAICS MEASUREMENTS

In order to obtain maximum possible energy from photovoltaic panel some of the prerequisites have to be met. One of such prerequisite is to consider the optimal angle and orientation of photovoltaic panel, which is often overlooked during installation and leads to lower energy gain as expected.

The electricity which is produced by photovoltaic panel is highly affected by its orientation and angle. In ideal state, when direct solar radiation is considered the optimal angle is equal to the latitude of the geographical location. Such prediction is not real, because the major part of the radiation is diffused radiation from other direction than the sun. Because of this, the optimal angle is slightly towards the horizontal. Season of the year must be considered as well, because it is obvious that in winter he sun sun´s elevation is low, so the panel needs higher angle in order to produce more energy and in summer the smaller angle of tilt is needed. The optimal orientation must be determined individually for each application. Local weather conditions such as morning fogs must be taken into account. Following figure shows the percentage of optimal energy production that can be expected from photovoltaic panel at different angles.(Fig.3)
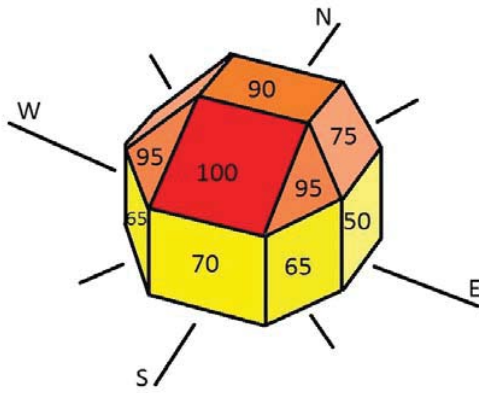
Fig.3 Percentage of Optimal Energy Production at Different Orientations and Angles

III. PHOTOVOLTAICS MEASUREMENTS DONE ON DEPARTMENT

Fully automated measurements of voltage and power of photovoltaic panel were done on KTEEM department. These measurements were done during whole year every day every minute and provide us with valuable data for further research.

The yearly overview of the panel´s average values of daily power (fig.4) and energy produced (fig.5) can be seen on pictures.
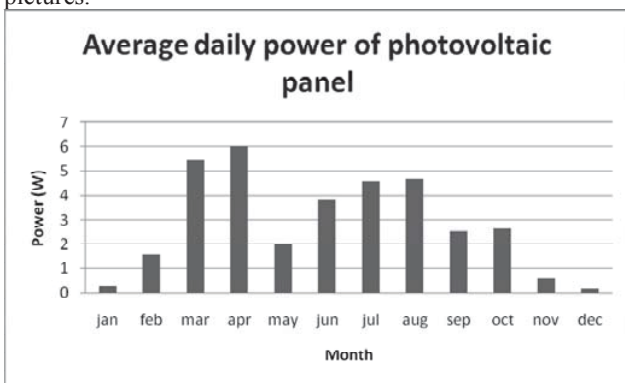
Fig.4 Average daily power of photovoltaic panel
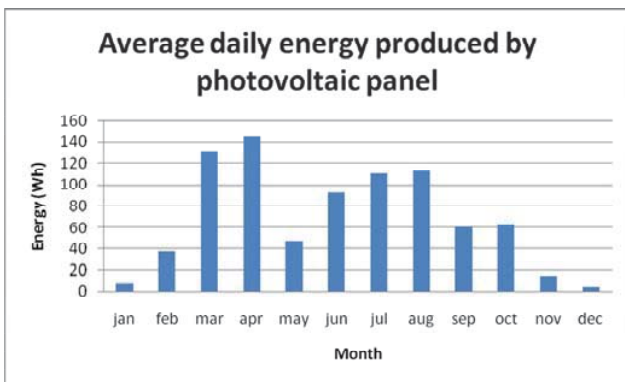
Fig.5 Average daily energy produced by photovoltaic panel

As we can see on the results, the photovoltaic energy is hard to predict, especially last year´s May was completely rainy, with

the weather absolutely unsuitable for photovoltaics use.

IV. PRICE OF PHOTOVOLTAICS

The photovoltaics technology is rapidly evolving these days, it is becoming more affordable, however its prices are still higher than the prices of traditional fossil fuels. However as can be seen on following picure (fig.6) the prices of the photovoltaics are decreasing. The prices are from year 2010.

Fig.6 Photovoltaic energy prices

V. OBJECT´S ENERGY DEMANDS

Under the term „object" we understand building which we want to evaluate. For the purpose of evaluation, program was written in MS Excel utilizing the embedded Visual Basic for programming the essential functions and calculation.

The object can be up to 2 floors building, where every storey can have from 1 to 15 rooms. Every wall can be specified (its size, construction material, type of plaster and/or insulation, also additional features like windows and/or doors).

From electrical point of view, the storey can be imagined in following way (Fig.7)
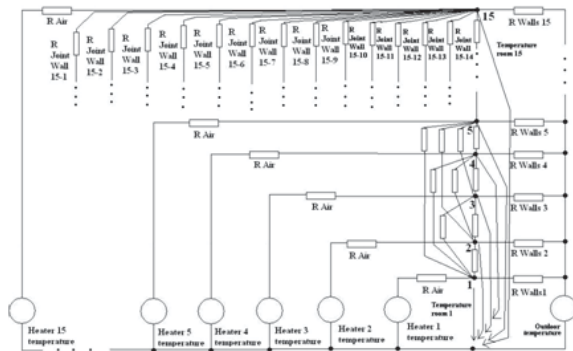
Fig.7 Simplified scheme of one storey

The drawing of this scheme is very simplified, as at the beginning we consider that all the rooms are connected with each other. In terms of logic it is of course nonsense, however the not defined joints would not interfere with the equations, as there will be zero value.

The resistors represent the total thermal resistance of walls (including any extra options like windows, doors, as well as ceiling and floor). The second resistor represents the air resistance. Every room has two sources, one is common to all, this is the outdoor temperature, which has to be specified. The second source represents the temperature of heater which will be counted after you specify the needed temperature in every room, which was specified. Needed thermal power output is counted as well.

Such scheme describes every storey, however in case of more storey, than just ground floor, the scheme includes additional resistors, which describe the thermal resistance of combined ceiling/floor and all of its features.

The system is as modular as possible, it is easy to add more storeys if needed, or even more rooms.
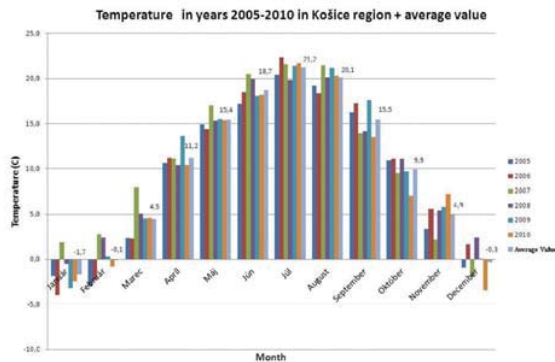
Fig.8 Temperature statistics for model

To obtain correct values for thermal power output, we must specify as most accurate values as possible. The thermal properties of construction materials are included in program, however for having a complete results, we must use the measured temperature values, ideally from whole year. Average temperature values from 2005-2010 years from Košice´s region will be used (Fig.8).

Fig.9 Sample screen of assessment program

As a reference object, we have considered the older residential house with one additional storey above the ground floor, which was made of older type full burnt brick, as it was the most common material, from which most houses are made from. The ground floor as well as the upper storey contain 4 rooms, so the total number of rooms is 8, which is quite standard. The outside temperature was chosen from the graph (Fig.8) and the inner temperature was set to 20 ºC to represent approximately the most common situation. We didn't consider any insulation, plaster or flooring materials, although it is possible, and needed for more precise results. The results in graphical form using MS Excel can be seen in the following picture (Fig.10)
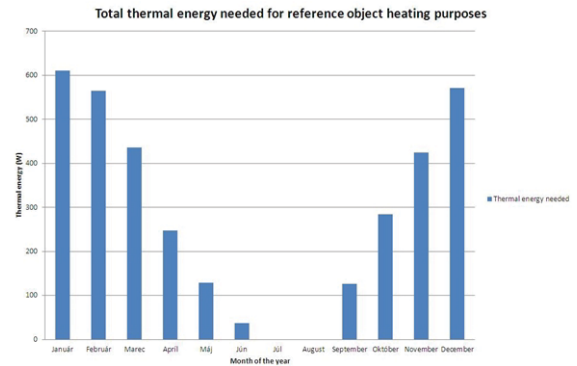
Fig.10 Thermal power output needed for object

VI. CONCLUSION

The results obtained with this assessment program will lead to next stage of research, which will be the creation of decision algorithm, which will be able to determine how to meet these energy demands using only the renewable energy sources in the most efficient way. The photovoltaics proves to be valid source, however it also proves to be very unreliable as a primary source of energy, therefore further research will be needed. It should be noted, that similar assessments of energy demand (for example electrical energy) can be done for wide range of applications, like for example automotive electronics.

REFERENCES

[1] Gaiddon Bruno, Kaan Henk, Munro Donna: Photovoltaics in the Urban Environment, Earthscan, 2009 England
[2] Green Martin: Third Generation Photovoltaics, Springer-Verlag Berlin Heidelberg, 2006Netherlands,
[3] Eicker Ursula: Solar Technologies for Buildings, John Wiley & Sons, 2003, England
[4] Petrova-Koch Vesselinka, Hezel Rudolf, Goetzberger Adolf High-Efficient Low-Cost Photovoltaics, Springer-Verlag Berlin Heidelberg, 2009, Germany
[5] Freris Leon, Infield David: Renewable Energy in Power Systems, 2008 West Sussex,

[6]   Molnar Jan: Remote Measurement System in Automobile Scheme, proceedings of OWD 2006, October 2006, Wisla
[7]   Molnar Jan: Telemetric System Based on Internet, proceedings of OWD 2009, October 2009, Wisla
[8]   Vince Tibor, Kovacova Irena: Distance Control of Mechatronic Systems via Internet, Acta Electrotechnica et Informatica, 2007
[9]   Vince Tibor: Motor Speed Regulation via Internet and Artificial Neural Network, proceedings of SCYR 2009, May, 2009, Kosice
[10]   Kovac Dobroslav, Kovacova Irena: Non-harmonic Power Measuring, Acta Electrotechnica et Informatica 3/2008
[11]   Kovac Dobroslav, Kovacova Irena: The Magnetic Fields of Electric Motors and their EMC, Advances in Electrical and Electronic Engineering, 1-2/2008
[12]   Kovac Dobroslav, Kolla Igor: Communication Interfaces for Measuring Systems, Acta Electrotechnica et Informatica, 4/2007

14

Possibilities of power flows control in electric power systems

[1]*Stanislav KUŠNÍR,* [2]*Vladimír KRIŠTOF*

[1]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
[2]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]stanislav.kusnir@tuke.sk, [2]vladimir.kristof@tuke.sk

Abstract— Electricity market opening has brought on the one side, a number of advantages and opportunities, but on the other side the number of problems. Lines are often overloaded and used to their limits, creating bottlenecks. This problem affects most countries, including Slovakia. Since building new lines is very tedious process, is necessary to introduce special equipments design to control power flow, such as PST, UPFC or FACTS.

Keywords— **FACTS, UPFC, HVDC, Power System**

I. INTRODUCTION

Opening of an electricity market brought a much of advantages, but on the other side, also much of drawbacks. Since commercially negotiated power flows are significantly different from the actual flows of trade laws and do not exceed the laws of physics, the question arises how to bring near these laws. As a result, of electricity trading has been increasing interstate transfers, often because of what some of the lines to the state, which are surcharge, while others are not fully utilized.

With reference to the market liberalization in the power industry, the actual tendencies of the electric power systems operation have the following character:

- Increasing capacities of the electric energy international exchanges in term of exportation from the sources, as well as in term of powers transit,
- increasing operating exploitation of the transmission elements, mainly international lines of interconnected electric power systems. This advanced form of the using of elements causes less reserves in case of the line surcharge. In background, the international lines were used mainly to increase the operation electric power system dependability in given area. Nowadays is overrides big business using, at what networks and their interconnections were not conceptually constructed,
- increasing differences between physical and business electric energy flows with the negative consequence to loss. These differences are still more expanding in last years,
- it exists relatively big unstableness and time changes of the transmission size. These processes are not possible to well predict,

- the networks operation is often adapted (by non-standard solutions too) to the business events,
- on these conditions, in any cases during the operation are beginning to detect networks bottlenecks, which can be limiting factor for the desired business changes. Sequentially, these situations can to cause the risk of the fail and breaking of the electric energy supply in areas.

The classic solution of the networks development (networks bottlenecks elimination), relative with the reinforcing and building of the new lines, is no wear as sufficient and quick in continuity with the problems to obtain of the new corridors and in connection with environmental problems. Therefore, often sought the ways, which could to enable at least regionally affect the negative functioning of therein before present processes for the transmission networks operation.

One of the decision distributive companies into the future is production of renewable for each household. Such production may result in rotation of direction load flow. Actual situation is possible to change and load flow will not from the transmission system to distribution, but on the contrary.

To arrive to effective operation of lines and prevent congestion interstate lines we can use devices designed to power flows control. One of the practically usable devices, which can be used to electric power system control in interconnected power systems are flexible alternating current transmission system (FACTS). [1], [3]

II. HVDC

HVDC - Hight Voltage Direct Current is an interface in the AC system, composed of rectifier, DC line, inverter and a range of accessories, especially filter to suppress the influence of higher harmonics on the ac systems. Using of HVDC technology is cost effective, but requires complicated and expensive devices on both ends of power lines. Therefore HVDC has found its application in the less conventional uses such as transmission of electricity on long distances, or connection of divided power systems, which can´t be operated on the base of traditional principles of AC systems. HVDC substations are very good regulated and allow quick change of size and direction of transmitted power.

GTW (Gas to Wire)

Gas-to-Wire (GTW) is a means of transmitting gas energy, as with a gas pipeline or the case of LNG and/or GTL.

GTW is for practical power generation adjacent to an oil/gas field, with optimum transmission by HVDC for AC-grid connection. [2]
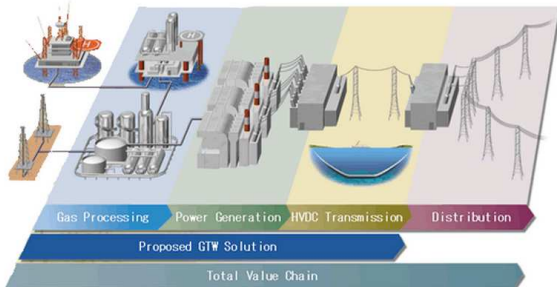
Fig. 1. Gas-to-Wire [2]

III. FACTS

Flexible (AC) electric transmission systems belongs to progressive technology in power system engineering. FACTS devices are used to optimize the existing transmission lines.

The term 'FACTS' covers all of the power electronics based systems used in AC power transmission [4].

The main systems are:

- Static var compensator (SVC)
- Fixed and thyristor-controlled series capacitor (TCSC)
- Phase-shifting transformer (PST) and assisted PST (APST)
- Synchronous static compensator (STATCOM)
- Synchronous static series compensator (SSSC)
- Universal power flow controller (UPFC)

Among adjustable parameters of these devices belong e.g.: voltage, current, impedance, phase angle.

A. TCSC

Principle of operation

TCSC configurations comprise controlled reactors in parallel with sections of a capacitor bank. This combination allows smooth control of the fundamental frequency capacitive reactance over a wide range. The capacitor bank for each phase is mounted on a platform to ensure full insulation to ground. The valve contains a string of series-connected high-power thyristors. The inductor is of the air-core type. A metal-oxide varistor (MOV) is connected across the capacitor to prevent overvoltages [4].
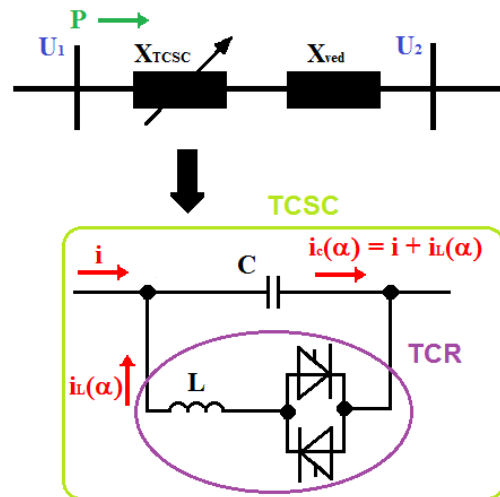
Fig. 2. Fixed and thyristor-controlled series capacitor (TCSC)

B. UPFC

Principle of operation

Converter 2 performs the main function of the UPFC by injecting, via a series transformer, an AC voltage with controllable magnitude and phase angle in series with the transmission line. The basic function of converter 1 is to supply or absorb the real power demanded by converter 2 at the common DC link. It can also generate or absorb controllable reactive power and provide independent shunt reactive compensation for the line. Converter 2 supplies or absorbs the required reactive power locally and exchanges the active power as a result of the series injection voltage. [4]

This paper is aimed at Unified Power Flow Controller UPFC.
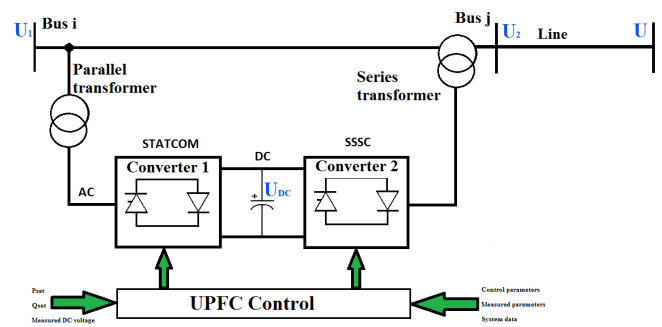
Fig. 3. Universal power flow controller (UPFC)
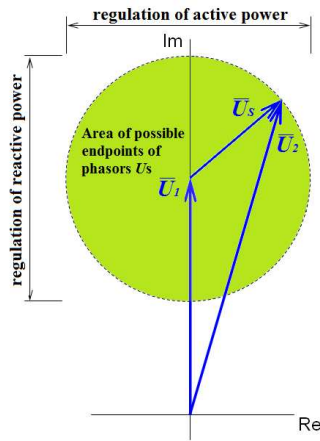
Fig. 4. Phasor diagram of UPFC

IV. ECONOMIC PROS AND CONS

Benefits of FACTS devices from economic perspective:

Building new lines is very tedious process (property settlement, land acquisition and etc.), is necessary to introduce special equipments design to control power flow.

Cons FACTS devices from an economic perspective:

- include semiconductor devices
- production is for order

These facts have an impact on their high price.

FACTS devices prices are approximately:

UPFC: 90 – 130 tis. US$/MVA
TCSC: 50 – 70 tis. US$/MVA

V. SIMULATION MODELS

A. 3 bus network

Easy 3 bus network is tested with and without UPFC inserted into two buses to analyze parameters (power flows and voltages) in this network.
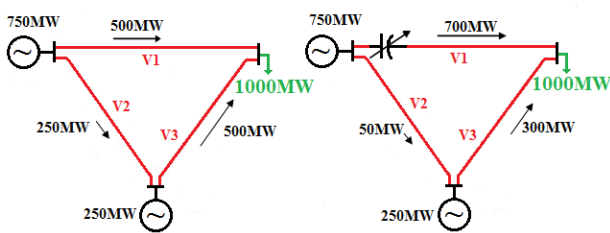
Fig. 5. 3 bus network
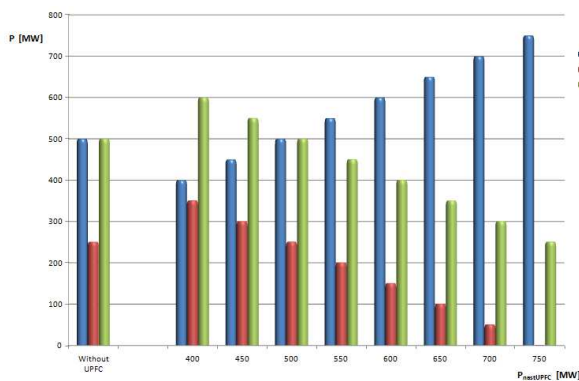
Fig. 6. Power flow control

B. 14 bus test network

Standard 14 bus test network is tested with and without UPFC inserted into two areas (13,8 kV and 69 kV) to analyze parameters (power flows and voltages) in this network.
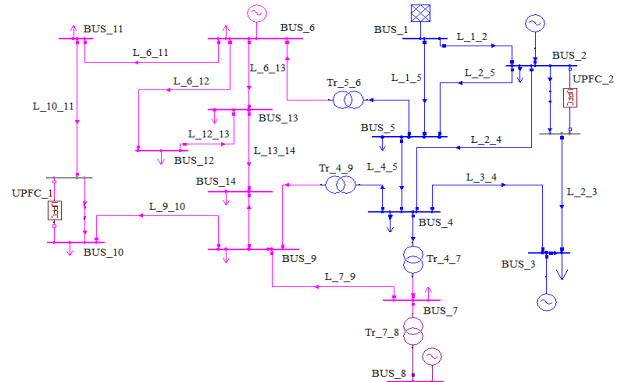
Fig. 7. 14 nodes network

UPFC_1 in Area 13,8kV

The reactive power flow of UPFC_1 is setting as constant. The UPFC_1 inserted into area 13,8 kV affects on the power flows mainly in this area. The change of power flows in area 69 kV is small.

The change of the power flow on the UPFC_1 causes the increasing of the active power losses in the network Fig.8.
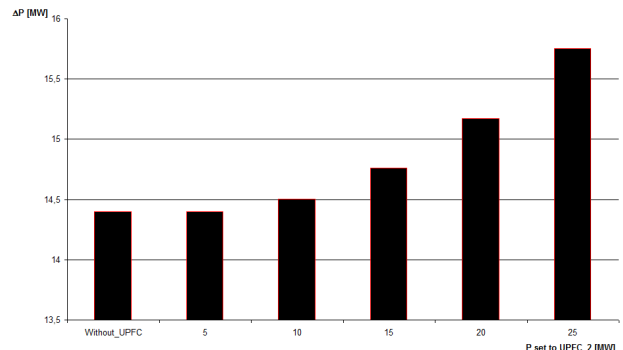
Fig. 8. Losses according to P set to UPFC_1

UPFC_2 in Area 69kV

The reactive power flow of UPFC_2 is setting as constant. The UPFC_2 inserted into area 69 kV affects on the power flows mainly in this area. The change of power flows in area 13,8 kV is small.

C. Power system of the Slovak Republic

Further research will focus to use FACTS devices in Slovakia. To solving of this issue, is necessary to model the power system of the Slovak Republic. However, as has been found to achieve the required accuracy is not sufficient only power system of the Slovak Republic, is necessary to model almost all systems, almost all belonging ENTSO-E. To the most accurate results is necessary to place the balance sheet of

the node somewhere outside of the Slovak Republic. Model of Slovak´s power system, has been successfully created and is fully operational. For the remaining models are constantly working.
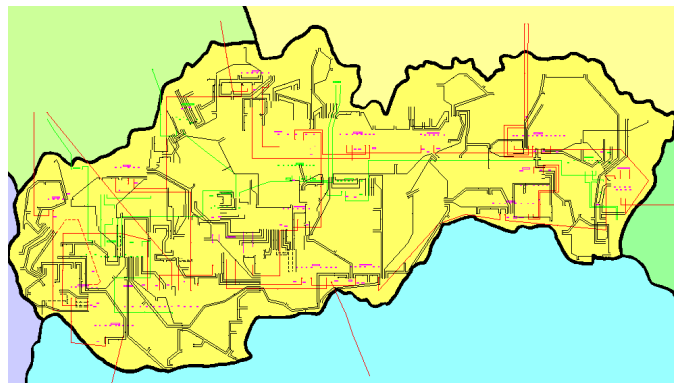
Fig. 9. Power System of The Slovak Republic

VI. CONCLUSION

UPFC belongs to the most integrated devices of the FACTS devices group. It is possible to use these devices for the in depended active and reactive power flows controlling. Its utilization in the electric power system control corresponds to their technical and economic demand. Currently, the investment cost not allows the mass installation of UPFC to the electric power system.

Gradually, however, will needs to introduce these devices more and more. Their benefits are not only in the regulation of power flows but help to improve the static and dynamic stability, oscillation damping and regulate voltage conditions. So we can assume that Slovak Power system over time cannot be without these devices.

ACKNOWLEDGMENT

REFERENCES

14

[1] Beňa, Ľ.: Využitie špecializovaných zariadení na reguláciu tokov činných výkonov v elektrizačných sústavách. Habilitačná práca. Košice: Technická univerzita v Košiciach, 2010, 84 s.
[2] GTW (Gas to Wire). [online]. Available on the Internet: <http://www.power-hitachi.com/products/ffp/pf/gtw.html.>
[3] Hlubeň, D.: Využitie transformátorov PST na riadenie tokov v ES SR. Dizertačná práca. Košice: Technická univerzita v Košiciach, 2009, 110 s.
[4] FACTS – Powerful systems for flexible power transmission. [online]. <http://www05.abb.com/global/scot/scot221.nsf/veritydisplay/b0f2c8c9 4b48a6bcc1256fda003b4d42/$file/facts_%20eng.%20abb%20review.p df.>

Post-processing the Output from Capacitive Senzor by Using Bandpass Sigma-Delta ADC Implemented on PSoC and MSI Circuit

Martin SEKERÁK, Marek GODLA

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

martin.sekerak@tuke.sk, marek.godla@tuke.sk

Abstract— **This article deals with realization of ΣΔ modulator, which is core block of bandpass ΣΔ Analog to Digital Converters are very often utilized in software determined communication equipment, software radio supported by manufacturer and individuals on utilization of programmable converters of narrow-band digital modulating signal. ΣΔ converters are suitable for implementation on VLSI technology because requirement the least possible amount of analog function blocks of A/D converter. Properties of this architecture of A/D converter led to realization of ideas to design modulator available on circuit System On Chip (SoC). Our used system SoC was developed by company CYPRESS MICROSYSTEMS, and specifically it is circuit CY8C27443-24PVI.**

Keywords— **Sigma-Delta converter, Analog-Digital Converter, PCoC.**

I. INTRODUCTION

Analog to digital converters (ADC) are today extremely widely utilize electronic block in various areas of techniques e.g. in communication systems [5], [6], computers [7], sensor techniques [1], [10] and elsewhere. There is large amount of the ADC architectures with large variety properties and parameters [8], [9] and another question of study is their testing [12], [13]. ADC are nowadays more and more used even in high-frequency communication systems and equipments like e.g. software radio, where very important part of them is bandpass sigma -delta ADC [5], [6].

The structure bandpass ΣΔ converter (BP ΣΔ) is based on the idea of LP ΣΔ modulator suppressing quantization noise around converting signal frequency, with possibility to adapt trade between converting resolution and sampling period by the digital processing algorithm in the output low-pass filtering. Substitution lowpass filter for resonator will cause, that lowpass modulator is change on bandpass. Resonator is realized by the time discrete structure in the surrounding of resonation frequency f_{IN}, which is determined by sampling frequency $f_S = 4.f_{IN}$. Besides positive properties like stability and dynamics, the noise transfer function (NTF) moves its minimum around the sampling frequency $f_S = 4.f_{IN}$ [5]. Binary multiplication with the reference signal f_{IN} shifts the spectral components at the modulator´s output around the DC component. Lowpass digital filter provides conversion of this component in the binary value. Phase shift of reference signal about one period Ts, provides digital value at the LP filter representing input signal component f_{IN} shifted by $\pi/2$.

The main benefit of BP ΣΔ converters is the direct conversion of the input analog signal to vector with two components (real and imaginary). High linearity caused by the LP filtering of the serial binary flux well knows advantage of ΣΔ converters.

Whole structure of BP ΣΔ A/D converter is shown on Fig.1. Resolution of each component at the digital output is adaptable by the length lowpass FIR filtration.
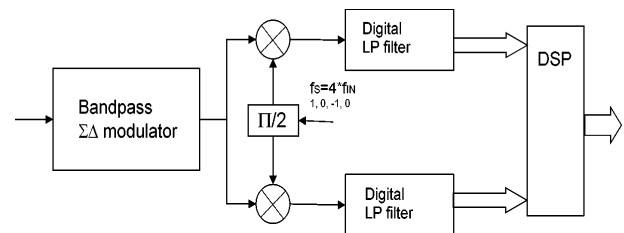
Fig. 1 Consequential AD converter for decoding I/Q component of harmonics signal.

These converters became very popular in the digital communications system using QAM, 16QAM, QPSK modulation like GSM, WCDMA, UMTS and by many next similar applications. Another prospective application of BP ΣΔ converters is sensor techniques [1]. Here the signal from the two-parameter sensor is gained, where one parameter affects the real component and the other parameter affects the imaginary component of the output impedance of the passive sensor. Alternatively, one component output impedance of the sensor can represent diagnostic information about the sensor functioning.
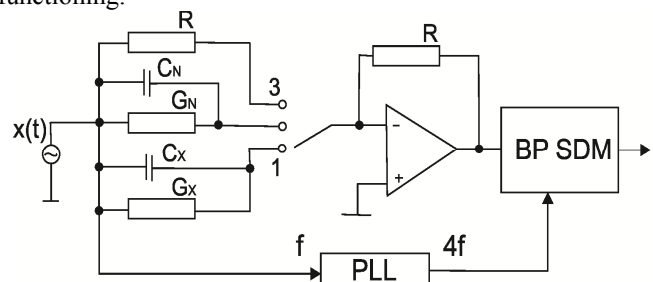
Fig. 2 Diagram of BP ΣΔ ADC as circuit for post-processing the output signal from the capacitive sensor

The phase shift of the processing block is suppressed by the subtracting operation. The difference of the digital values from the output of the LP filter in the position 1 and 2 normalized to the value measured in the position 3 is expressed by the formula. Here the measured changes of capacity and resistance are obtained:

$$\frac{U_1 - U_2}{U_3} = R\left(j2\pi f_s \Delta C_X + \Delta G_X\right) \qquad (1)$$

The BP $\Sigma\Delta$ A-D converters can be build-in using various mixed signal technologies. They contains minimal amount of analog elements (comparator, two integrators and summing circuit) and their offsets are compensated by the feedback structure. Simple analog structure causes high linearity.

II. PROGRAMMABLE SYSTEM ON CHIP (PSOC)

Circuit Programmable System On Chip (PSoC) of company Cypress Microsystems allows integrate analog digital programmable blocks into one integrated circuit. System On Chip (SoC) consists of hardware and software part, which is control microprocessor and peripheries. Used circuit CY8C27443-24PVI is created on the SONOS technology. Microprocessor called M8C is building on 8-bits hardware architecture and allows an effective programming. Digital part consists of FPGA blocks organized in rows. Analog blocks are used for development analog element like comparators, summing amplifier and AD and DA Converters. Configuration and emulation of circuit is being performed in the graphical development environment PSoC Designer.

Analog part contains three types of the analog blocks: Continuous Time (CT) provide continuous time analog function. Two other represented by the Switching Capacitor blocks (SC) of type C and D, provide programmable signal amplification or integration using switching capacitor technology as shown on Fig. 3. Blocks are organized in the columns, where each has possibility to connect its output into inputs of other ones. Besides interconnection of the blocks clock signal and the reference voltage could be connected to the inputs of analog blocks. State of the interconnection switches is being memorized in the mapping memory. This Flash memory contains information enabling or deactivating internal phase shifted clock signals and selection/activations of the input multiplexers.
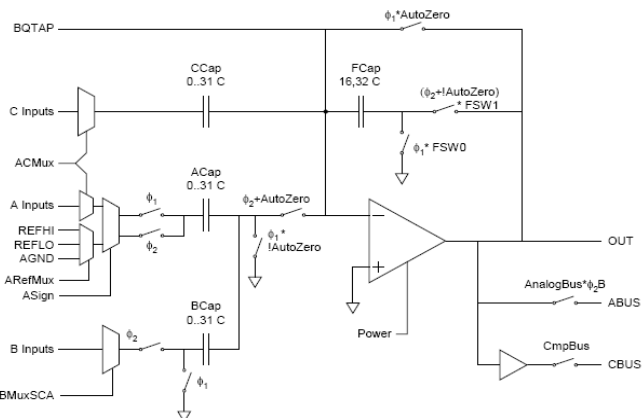
Fig. 3   Analog switched capacitor block type C

III. DRAFT BP SD MODULATOR CIRCUIT FOR PSOC

The resonator of $\Sigma\Delta$ modulator is designed by supplementing a double-delay circuit of the negative feedback circuit and summing circuit in the analog part of PSoC. The input analog signal is connected to the input amplifier together with the signal from the comparator´s output through a resistive summing circuit. The output signal then passes the double-delay resonator. The double-delay circuit is implemented by means of two SC blocks, each of them representing an amplifier with the amplification equal to one, and creating delay of one f_S clock. The transfer characteristics of the double-delay circuits with the used unity amplification are shown in Fig. 4. The effect of the SC block saturation deteriorates characteristics of higher input amplitudes.
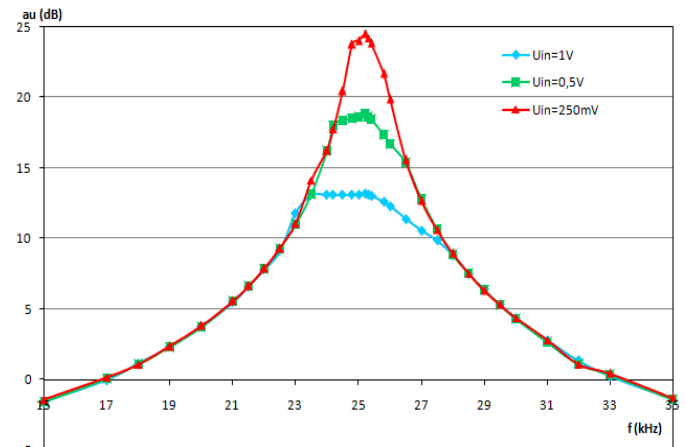
Fig. 4   Transfer characteristics of a resonator

The comparator is connected to the resonator´s output. The comparator output is a digital bus connected to the buffer located in the digital circuit. The configuring BP $\Sigma\Delta$ modulator in analog parts of the PSoC circuit created in the graphical development environment is shown in Fig. 5.

The input signal range is restricted by supply voltage of 0 - 5V. In order to remove unilateral restriction, the DC bias of 2.5V was added to the input signal.
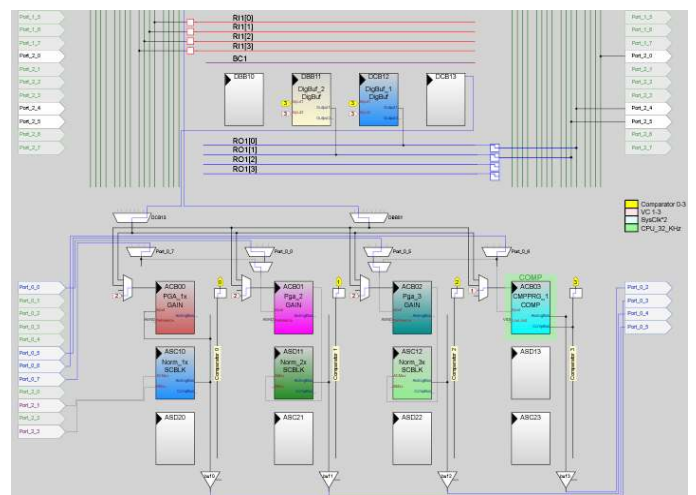
Fig. 5   Diagram BP $\Sigma\Delta$ modulator in development environment PSoC Designer

IV. IMPLEMENTATION OF BP ΣΔ CONVERTER USING DISCRETE COMPONENTS

Limited amount of the analog blocks for layout modification towards symmetrical structure, small utilizable input signal range with the required DC offset are drawback of the implementation on SoC blocks. The BP ΣΔ modulator was also implemented using MSI components.

The delay-block z-1 was realized by a pair of sampling amplifiers (SA). The non-overlapping clock with 50% duty cycle on each SA in one block ensures time delay z-1. Comparators and amplifier were implemented on broadband integrated circuits. Resistive summing circuit maintains an accurate summing operation without drift. Final circuit structure of BP ΣΔ modulator implemented on MSI circuits is shown on Fig.6.
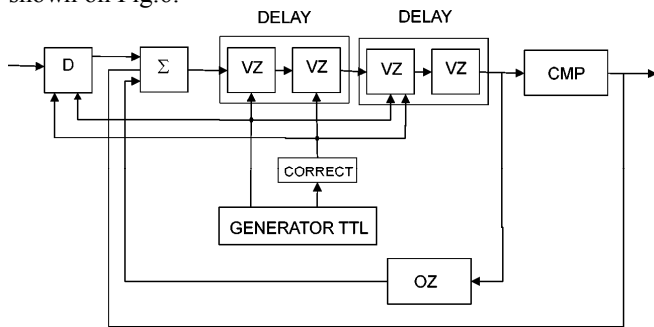
Fig. 6   Structure bandpass ΣΔ modulator realized using MSI circuits

Acquisition of the binary output from the BP ΣΔ modulator was performed by DAQ NI USB-6521 for both circuit implementations. Initially the acquired binary flux from the BP ΣΔ modulator was multiplied by two sequences of the binary sampling signal mutually shifted by one tact. This sampling shift corresponds to the phase shift π/2 of the input signal f$_{IN}$. The successive signal LP filtering was performed by a hosting PC, where the program was created in LabView development environment by NI.

V. EXPERIMENTAL RESULTS

The measurement both implementations of the implemented structures BPΣΔ modulator was performed by the oscilloscope, and simultaneously by the DAQ card and registration program for binary data from the modulator created in LabView.
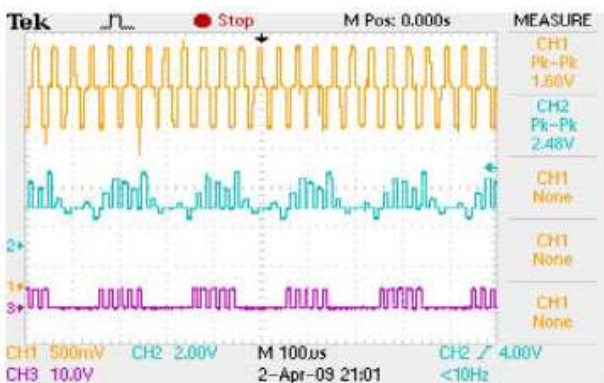
Fig. 7   The measured course of the BP ΔΣ modulator output from the oscilloscope for the input signal amplitude Uin=0.5V

The measured course of BP ΣΔ modulator on a single value of the input voltage U$_{IN}$=0,5V on oscilloscope record is shown in Fig. 7, while the resulting binary sequence is in Fig. 8.
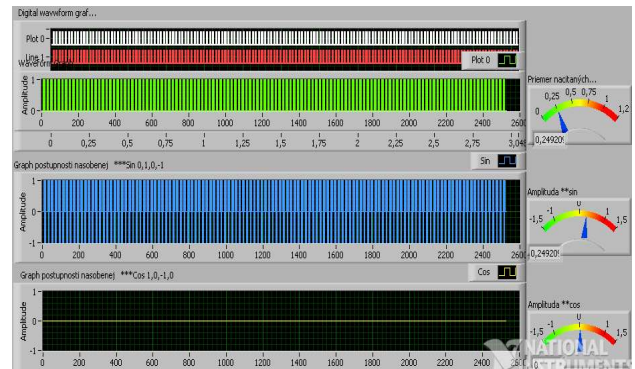
Fig. 8   Course and numerical evaluation of the binary data from the BP ΣΔ modulator output for amplitude of input signal U$_{IN}$ =0.5V obtained by means of DAQ card and the binary data registration program.

The multiplication by the binary sampling frequency moved the noise shaped signal from BP ΣΔ modulator to DC at the frequency axis. The resulting signal is filtered by a digital low-pass filter. The transfer characteristics for SoC as a function of the feedback resistance can be seen in Fig. 9. The linearity of the transfer function between the input signal amplitude and digital output is limited. The slope of the transfer characteristics and the input signal range can be controlled by the resistance in the feedback of BP ΣΔ modulator.
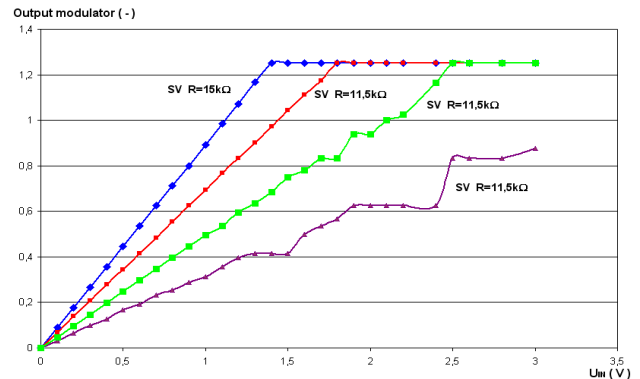
Fig. 9   Resulting transfer characteristics of BP ΣΔ modulator realized on PSoC circuit

One of the major characteristics of this converter is its sensitivity to phase of input signal, as reflected by changing values of I and Q components obtain from the processing of binary information from modulator. Fig. 10 shows the dependence I and Q components on the phase of the input analog signal to be converted.
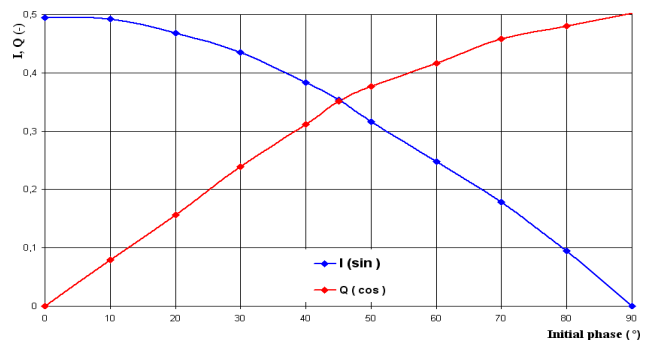
Fig. 10   Dependency of I and Q components from the initial phase of analog signal

VI. CONCLUSION

$\Sigma\Delta$ modulators retains many advantages compared with other types of converters, which include simple feasibility since they are based on digital processing in favor of minimal amount of analog blocks. SoC belongs to the group mixed signal circuits where both digital and analog part is programmable according to user´s needs. Especially digital multiplication and low pass filtering can perform on the same chip as analog preprocessing operations. The main shortcoming of this solution is small range of linearity of the transfer characteristic and restricted possibility to perform analog operation in the differential mode. Additional offset causes thermal drifts in the analog preprocessing blocks.

Implementation of $\Sigma\Delta$ modulator on MSI components reduce this drawback paid by no consistent block structure with both analog and digital integrated circuits. The strength of this implementation is possibility to modify any part independently. The most prospective are ASCI circuits for user with available design tools.

REFERENCES

[1] Haze, J; Fujcik, L; Vrba, R, Michaeli.L. et al.: "The Utilization of Novel Bandpass Sigma-delta Modulator for Capacitance Pressure Sensor Signal Processing" , Radioengineering Volume: 17 Issue: 4 Pages: 79-85 Published: 2008

[2] Krukowski A., Kale I., Morling R. : "Applications of polyphase filters for bandpass Sigma-Delta analog-to-digital conversion", International Workshop on ADC Modeling, May 1996

[3] C. I. Lao, H. L. Leong, K. F. Au, K. H. M and S. P. U, "A 10.7-MHz bandpass sigma-delta modulator using double-delay single-opamp SC Resonator with double-sampling", IEEE Circuit and System, May 2003, pp 1061-1064.

[4] Brigati S., Francesconi F., Malcovati P., Tonietto A., Maloberti F.: "Modeling Sigma-Delta Modulators Non-Idealities in SIMULINK", Proceedings of IEEE International Symposium on Circuit and Systems, Orlando USA, 1999

[5] Salo T., "An 80-MHz bandpass $\Delta\Sigma$ modulator for a 100-MHz IF receiver", Solid-State Circuits, IEEE Journal of, vol. 37, pp. 798-808, 2002.

[6] H. K. C. Issac, "A 70 MHz CMOS band-pass sigma-delta analog-to-digital converter for wireless receivers", M.S. thesis, Hong Kong University of Science and Technology, August 1999

[7] ŠALIGA, Ján - MICHAELI, Linus: Software for metrological characterisation of PC sound cards. In: Computer standards and interfaces. vol. 25, no. 1 (2003), p. 45-55

[8] MICHAELI, Linus - ŠALIGA, Ján - KOLLÁR, Martin: Parameters of band pass sigma delta ADC and the comparison with the standard ones. In: Measurement. vol. 40, no. 5 (2007), p. 473-478.

[9] ŠALIGA, Ján: Testovanie parametrov analógovo-číslicových prevodníkov dynamickými metódami. In: Slaboproudý obzor. roč. 59, č. 1 (2002), s. p6-p11.

[10] KOLLÁR, Martin - ŠALIGA, Ján: The principle of new sigma delta modulation technique based upon the use of a flip-flop. In: 13th International Symposium on Measurements for Research and Industry Applications: Athens - Greece, 2004. ISBN 960-254-644-1.

[11] Michaeli, L; Saliga, J; Kollar, M.: "Parameters of band pass Sigma Delta-ADC and the comparison with the standard ones", MEASUREMENT Volume: 40 Issue: 5 Pages: 473-478 Published: 2007

[12] MICHAELI, Linus - MICHALKO, Peter - ŠALIGA, Ján: A new ADC fast testing method based on the unified error model. In: Measurement. vol. 41, no. 2 (2008), p. 192-197.

[13] SERRA, Antonio Cruz et al.: Combined spectral and histogram analysis for fast ADC testing. In: IEEE transactions on instrumentation and measurement. vol. 54, no. 4 (2005), p. 1617-1623

14

Real-Time Regulation Systems Based On Internet Integration Algorithm Into The Existing Network

Tibor VINCE

Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

tibor.vince@tuke.sk

Abstract—**The article presents one of the ways for implementing Internet as a communication bus between real-time system elements. The paper describes algorithm, which helps to build network for a real-time system. This algorithm is suitable for creating new network, or own domain of system elements. For implementing system in existing network, there would be few modifications in algorithm.**

Keywords—**algorithm, regulating system, Internet, remote control**

I. INTRODUCTION

There is huge effort to integrate different cooperating systems in one complex system. The basic problem is communication between these different modules of the system, especially when the modules are located in different locations. According to communication requirements, appropriate communication way has to be chosen.

The Internet plays an important role not only in information retrieving, but also in industrial processes manipulation. It is difficult to build a real-time control network using the standard Ethernet because the Ethernet MAC (Medium Access Control) protocol - persistent CSMA/CD (Carrier Sense Multiple Access with Collision Detection) protocol may cause unpredictable access delay.

Requirements put on Internet depend on the system. Since each system has its own requirements, there is no prototype of such Internet bus. Therefore there should be an algorithm for building or implementing such suitable network. Specific algorithm depends on the number of system elements (if system is complex or not), if system should be implemented in existing network, or it is necessary build a new one etc. This contribution solves the cases, when regulating system needs to be implemented in to the existing computer network.

II. ANALYSIS AND PROPOSALS

The whole diagram is shown in figure Fig. 1. First step is dataflow analysis. The most important question that needs to be answered is the following: What amount of data will be transported over the network and what is the greatest acceptable time delay? The next step is acquiring statistical data of existing network. It has to be taken into the consideration that network delay is variable so the delay measurements should be processed long time or multiple measurements in different time. The most important are average delay and maximum delay peaks. Then the measured data may be compared with data flow analyses. Measured statistical data may be now compared with data flow analyses. If the system requirements are met, it is possible to apply maintain algorithm shown in figure Fig. 2. When the regulating system is implemented into the existing network, there is assumption that other users or systems have already used the network. Even the requirements are met even in the delay peeks in present, it is not possible to exclude situation that the requirements will not be meet in the future. Therefore system has to be prepared for such situation.

III. NETWORK IMPROVE METHODS

If the network doesn't meet system requirements, it should be improved. Among the ordinary methods of improvement we can include: improvement of network architecture, improvement of network bandwidth and special protocols implementation. The sequence of these methods in algorithm can vary depending for example on the cost of implementing these methods. Improvement of network architecture can lead to the increase of the number of nodes, branches and changing architecture.

The goal is to decrease network traffic jam and that way decrease packets time delay in. Classic Ethernet has 10MBps bandwidth and has reserve in 100MBps and also 1GBps bandwidth. Widening Ethernet bandwidth is another way to decrease packet delivery time. It is possible to decrease time delay by using special protocols, for example RSVP-Resource Reservation Protocol, IPv6-support packet priority.

Applied special protocols must be supported by all network elements. Special protocols are meaningful only in complex systems. Time deterministic Ethernet is a special way for decreasing time delay. For example: Industrial Ethernet, Ethernet Powerlink etc. Deterministic timing is achieved by applying a cyclic timing schedule for all connected nodes to access the physical layer. To apply deterministic type of Ethernet, all system elements in a real-time domain must support such type of Ethernet.

The last chance to use Internet is to modify system requirements. Sometimes, with small system changes, can modified system requirements meets the network parameters.

IV. ALGORITHM

The whole diagram of new network generation is shown in figure Fig. 1.
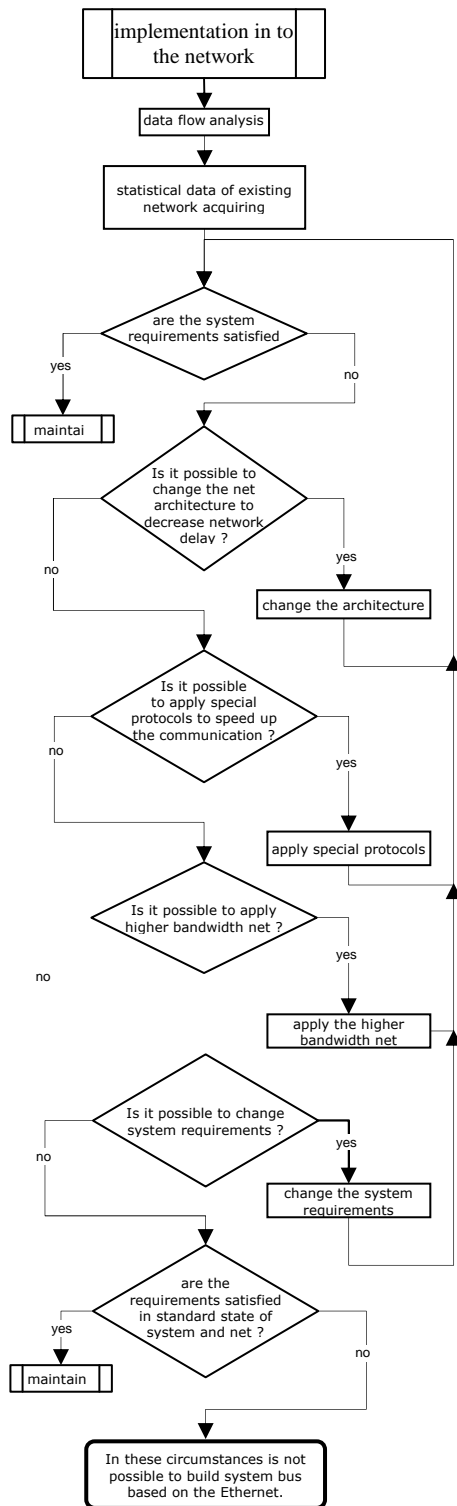
Fig 1. Implementation in to the network algorithm

There could be situation when Ethernet could not fulfill regulation system requirements. Very important information is, if the requirements aren't accomplished in standard situation or just in special occasion (with certain probability). If problems are standard situation, the remote regulation system could not be used in the existing network. In case of special situation the time delay exceeds the system requirements, the system needs to be maintained to handle

such situations. Such a maintain algorithm is shown in figure Fig. 2.
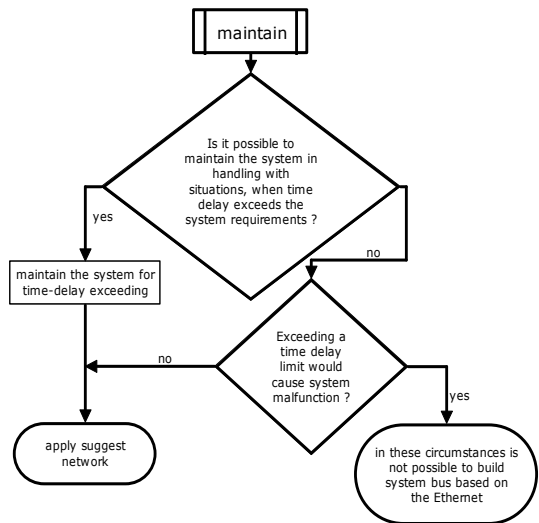
Fig 2. System maintain algorithm

V. CONCLUSION

These days, there is effort to utilize Internet also in real-time systems. There are already real-time systems using computer network but all problems were solved for specific system. There are not universal guides yet. This contribution is an attempt to fill this gap.

This algorithm describes one of the possible ways. The whole algorithm aims to build a network that is as simple as possible and gradually improve it until the network parameters meet the system requirements. Requirements of feedback less than tens of milliseconds can be achieved only with deterministic type of Ethernet, but all devices in real-time domain must support such type of Ethernet. T

REFERENCES

[1] Yang, S.H., Tan, L.S., Chen X: Requirements Specification and Architecture Design for Internet-based Control Systems, proceedings of the 26th Annual International Computer Software and Applications Conference (COMPSAC'02), 2002.

[2] Hall E.: Internet Core Protocols: The Definitive Guide, O'Reilly & Associates (February, 2000) USA, ISBN: 1-56592-572-6.

[3] http://www.anybus.com/technologies/technologies.shtml, 10.2.2008.

[4] Kováč, D., Kováčová, I., Molnár, J.: Electromagnetic Compatibility - measurement, TU Košice publisher, 72 pages, ISBN 978-80-553-0151-8.

[5] Fonda C., Postogna F.: Computer networking basics, ICTP workshop on telecommunications: science, technology and applications. Trieste, 15th Sep. - 3rd Oct. 1997

[6] Molnár, J., Kováčová, I.: Distance remote measurement of magnetic field. In: Acta Electrotechnica et Informatica, 2007, Vol.7, No.4, pp. 52-55, ISSN 1335-8243.

[7] Molnár, J.: Telemetric system based on Internet. In: OWD 2009: 11 International PhD Workshop: Wisa, 17-20 October 2009, p. 38-41. ISBN 83-922242-5-6.

[8] Kováč, D., Vince, T., Molnár, J., Kováčová, I.: Modern Internet Based Production Technology. In: New Trends in Technologies, SCIYO Publisher, Croatia, 2010, pp. 145-164, ISBN 978-953-307-212-8

[9] Kováč, D., Kováčová, I.: EMC Aspect as Important Parameter of New Technologies. In: New Trends in Technologies: Control, Management, Computational Intelligence and Network Systems, Publisher: Sciyo, November 2010, pp. 305-334, ISBN 978-953-307-213-5

ACKNOWLEDGMENT

14

Reliability Analysis for Transformer Feeder Protection

Marian HALAJ

Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

marian.halaj@centrum.sk

Abstract—**Submitted paper is focused on the reliable design of power system protection. One of the most effective ways to improve the reliability of protection system is an effective redundancy. The analysis method is described to improve the reliability of power system protection.**

Keywords— **Reliability, Security, Protective Relay**

I. INTRODUCTION

From the reliability and safety point of view the protective relays and circuit breakers are very important components of power distribution system. The appropriate protection system should recognize and selectively to trip any fault occurred. Such a function is normally ensured by means of protective relays redundancy, in a general way with the relays working on the different physical principle. However, much less attention is paid to the reliability analyses of the protection circuit itself. The overdesign of the protection system and related high concentration of the components may lead in different result than expected, e.g. to the lower reliability level. In addition to that, the optimization of power system protection may significantly save the initial and operating costs of the power system. The purpose of submitted paper is an introduction to the problem and comparison of the most common protection techniques on the model of distribution transformer.

II. PRINCIPLES OF TRANSFORMERS PROTECTION

The faults on the transformers can be divided into two main categories:

- Internal failures
- The faults caused due to current flow through transformer

The internal failures may be isolated using short circuit and earth fault relays. The most significant cases in this category may be considered the short circuits on the transformer terminal, on the winding and earth faults.
The failures caused by current flow through the transformer may be identified using overload protection. To protect the medium and large oil-filled transformers are usually protected by the following types of protective relays:

- Short circuit relay
- Earth fault relay
- Bucholz relay
- Differential protection relay

The differential protection is commonly considered as a most effective transformer short circuit protection. Because of very short reaction time (generally in tenths of ms) is very helpful in order to prevent the transformer from the occurrence and propagation of serious transformer failures. The application of this protection relay is unfortunately quite difficult and very expensive. Generally this type of protection is applied on the large or very important transformers.

Typical block diagram of modern numerical relay is shown on the Figure 1. In general, the protection system consists of the following components:

- Current and voltage transformers
- Input filters
- Binary Input/output boards
- CPU
- Tripping and closing coils
- Power supply
- Interconnecting cables and wires

One of the main advantages of modern numerical protective
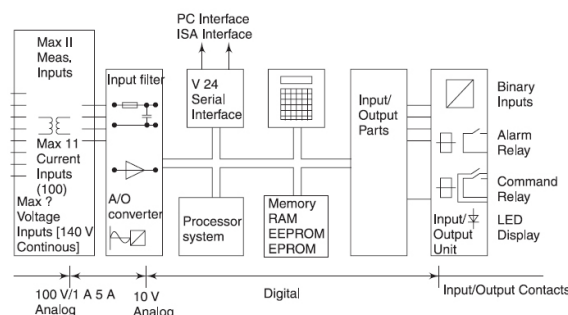
Fig. 1. Block diagram of numerical protective relay.

relays against the conventional relays is 'complexity'. The same hardware unit may be used to provide several protection functions against the separate relays needed in case of electro-mechanical relays used. In certain cases the back up protection can be provided using another tripping characteristic, i.e. the overcurrent protection function (ANSI 51) can trip the failure in case the differential protection (ANSI 87) doesn't do the same. In such cases the redundant converters and binary inputs are installed only and no additional relay is required.

III. RELIABILITY AND REDUNDANCY

Reliability is a product of two factors [1]; dependability and security. For relay system protection, dependability is defined as the ability to trip for a fault within its protective zone while security is the ability to refrain from tripping when there is no

fault in the protective zone (ability not to trip when not called for). As we can see these requirements on the protective relays are mutually exclusive. It means that the process of power protection optimization is to maximize the bivariant function.

One of the most effective ways to improve the dependability of protection system is redundancy. Redundancy is defined as 'the existence of more that one means for performing a given function [5]. It is obvious that protective relay system dependability can be increased by added redundancy as if one of the systems does not trip for an in-zone fault, a redundant system may. Security on the other hand, is generally decreased by increased redundancy as there are added devices in the system that may trip when not called upon to do so. However, redundancy does not influence dependability and security to the same degree.

IV. RELIABILITY EVALUATION

To establish a simulation model and perform the reliability evaluation the term 'failure' shall be defined. The word failure means the system state when the system is not able to perform the required functions. This analyze is limited to the failures caused by electric components malfunction. The failures on the communication system, software defects or wrong settings are excluded. Therefore the tast is mainly focused on the following situations:

- Malfunction of current and voltage transformers
- Short circuit or interruption of interconnecting cables
- Defects of protection input circuits
- Malfunction of A/D converters
- Loss of auxiliary voltage supply
- Failure of tripping coils (failures of closing coils is not considered as important failure)

In reference the above mentioned situations the following actions are considered as most effective to improve the entire system reliability:

- Independent auxiliary power supply from several batteries
- Redundant tripping circuits with different tripping coils
- Binary I/O redundancy
- Redundant current and voltage transformers
- Redundant A/D converters

The failure rate estimation has been made by use of a fault tree method. Results of the typical simulation are borrowed from [1]. Table 1 shows the calculated failure rate for 3 situations:

- Simple protection system without redundancy
- 100% redundant protection system
- Sophistic redundant protection system

The failure rate estimation has been made by use of a 'Fault Tree'. An AND gate represents the probability that both modules fail at the same time for the output to fail. An OR gate represents that if any module connected to this gate fails, the output will fail. Consequently, the failure rate for an AND gate is the product of the inputs while the failure rate of an OR gate is the sum of the inputs. All redundant modules in the system are connected via AND gates, and the outputs of these are finally combined in an OR gate. The redundant

Protection modules comprise both the input transformer circuitry and the microprocessor module. These two modules are therefore combined in an OR gate before entering the Protection AND gate.
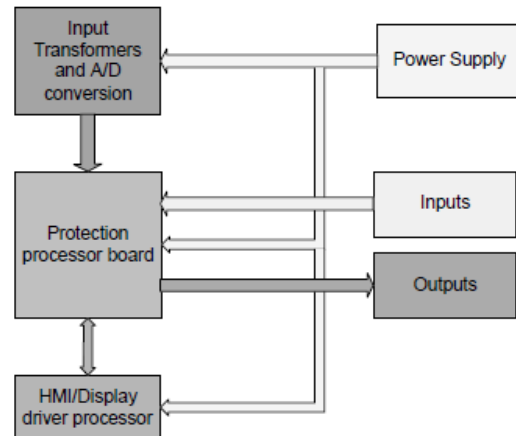
Fig. 2. Relay hardware block
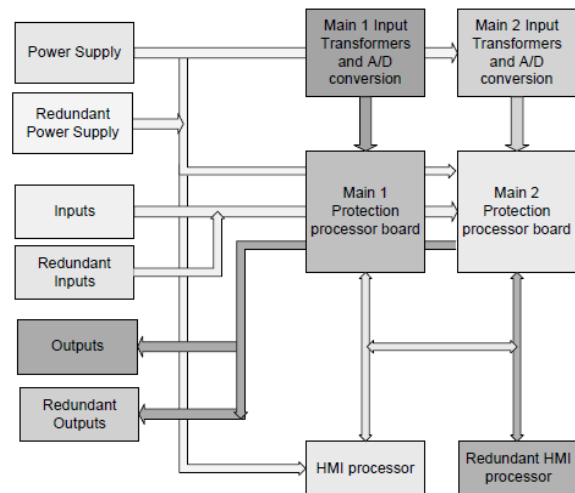
Fig. 3. Sophistic redundant protection system
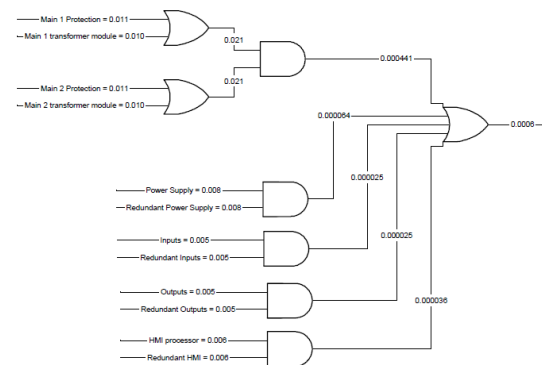
Fig. 4. Failure rate for sophistic redundant protection system

The results show the significant difference between the failure rates of different system configuration. Calculated failure rate for sophistic redundant protection system is approx. 3 times lower than failure rate of 100% redundant system although the investment costs of such a system are lower.

System	Failure Rate (failures per year)	MTBF (years)
Single system	0,045	22
Conventional redundant system	0,02	500
Sophistic redundant system (Refer to Fig. 4)	0,0006	1667

Table 1. Failure rate for sophistic redundant protection system

V. CONCLUSION

The results of the reliability analyses performed on the sophisticated redundant system show the improvement of the overall failure rate by factor 3 and more against the conventional redundant system. The calculations are performed on the basic power protection models only therefore the extended research work is going to be performed in near future to investigate the reliability of the real power distribution system protection including the interfaces to the substation control system in order to improve the failure rate of existing or newly built power distribution systems.

REFERENCES

[1] T. Dahlin, W. Higinbotham, "*Improving Reliability for Power System Protection*", 58ᵗʰ Annual Protective Relay Conference, Atlanta, April 28-30, 2004.

[2] F. Janíček, V. Chladný, A. Beláň, Ž. Eleschová, Digitálne ochrany v elektrizačnej sústave. Bratislava : STU, 2004, 360 s., ISBN 80-227-2135-2

[3] J. Tuma, S. Rusek, Z. Martínek, M. Tesařová, I. Chemišinec, Security, quality and reliability of electrical energy, Conte spol. s.r.o., ČVUT Praha, 2007, 172 s., ISBN 978-80-239-9056-0

[4] IEC 61078: 2006, Analysis techniques for dependability – Reliability block diagram and boolean methods.

[5] The IEEE Standard Dictionary of Electrical and Electronics Terms, IEEE Std.100-1996

14

Renewable Energy to Reduce Energy Costs for Industrial Park Kechnec

Matúš NÁROŽNÝ

Department Of Management and Economics, SjF TU of Košice, Slovak Republic

matus.narozny@tuke.sk

Abstract—**Solar energy, the clearest energy on Earth. It is an important source for our new generations and using of sources like energy of sun, should help us to save nature for next generations. I would like to present possibilities in using of solar energy at industrial parks and it is because they are the biggest buyers of energy ever.**

Keywords—**Solar energy, industrial park, energetic costs.**

I. Introduction

The idea of an industrial park in Kechnec came around year 1996. Village management systematically begun preparing conditions for potential investors who could bring new jobs to the region. With first investor - Molex - intensive preparation of the area began. In October 2003, the industrial park Kechnec was inaugurated. Today Kechnec offers not only real estate and infrastructure, but also necessary services related to health care, housing for staff, and sports and cultural activities.

II. Progress of the Construction of Industrial Park in Kechnec

In April 2002 the Government of the Slovak Republic (SR), supported the establishment of an industrial park Kechnec with a subsidy of 184.3 million Sk. From this subsidy new infrastructure was built, which reaches to the limits of the park and covers gas, water, electricity, sewerage, sewage treatment plant, access road and telephone lines. From the state subsidy also fees for permanent land occupation in the range of I. stages of the industrial park (80 ha) were and will be paid. Slovak Government approved the second state subsidy of 28.8 million in February 2004 and it is intended to finance infrastructure.

Almost two years passed since the opening of the industrial park. On the area of 5 ha 4 new investors: Dorsvet Plus, Evans, and Imrich Čamaj JISIMEX are building their plants. Kechnec village has signed a contract with one strategic investor,which is the company Getrag Ford Transmissions, and three smaller investors: Plastipak, Kuenz and VODS [1]. Currently Kechnec industrial park employs more than 1500 workers. The park development has also a positive impact on the increase of population in Kechnec.

Within two years, the village will be ready for construction of up to 200 flats. Primary and secondary health care in Kechnec is provided by professionally equipped clinic. There is also a sports center with two terragreen surfaces suitable for tennis and football, massages, sauna, various forms of rehabilitation. The whole complex has a professional lighting and sound equipment.

III. The Infrastructure of the Industrial Park

Kechnec Industrial Park is divided into three zones:

- 80 ha for small and medium enterprises,
- 200 ha for strategic investors,
- 52 ha for logistics center.

Currently, the total size of the industrial park Kechnec is 332 ha. The territory of the first stage of the industrial park (80 ha), which is now fully operational, is in the report of KIPP, a. s. - manager of the industrial park in cooperation with the municipality Kechnec (http://www.kechnec.sk). Built-up area for small and medium enterprises within the existing industrial park is 10 ha (Gilbos Slovakia, JISIMEX, Imrich Čamaj, Evans and Dorsvet Plus).

Contracts signed with investors are already for about 22 ha (Plastipak, Kuenz, VODS). The remaining approximately 50 ha is excpected by the community to attract investors who will create min. 3000 jobs and these investments should take into account environmental concerns and community use of community services offered [2]. The Getrag Ford Transmissions company project is considered a significant investment, implementation of which is in the public interest. The companyâĂŹs plant will therefore be situated in the area for major investors with a total area of 26 ha.

1) Industrial Park Kechnec:
   - 332 ha of free soil.
   - Newly constructed public infrastructure – gas, water, electricity, waste disposal, sewerage, telecommunications connections, driveway, sewage treatment plant [3].

2) Region Košice-ambient:
   - More than 23% unemployment.
   - The average monthly wage is around 300 EUR.
   - 400 000 people of working age live within a radius of 30 km.
   - At a distance of a one-day journey by truck investor has access to 350 million people there are markets of 13 countries within a radius of 1,000 km [3].

3) Transport links:
   - Industrial park Kechnec is 18 km from Košice the second biggest town in Slovakia International airport Košice is 18 km away.
   - Industrial Park is located 0.5 km from the border with Hungary. Entry into Ukraine and Poland is by road about 100 km.
   - Connection to the European motorway network via Hungary from the borders of the park is about 60

km, while in short term it is planned to finish the piece between Kosice (Slovakia) - Miskolc (Hungary). After joining the European Union, continuous highway was created from Kosice to Europe. The road I/68, which passes Kechnec, is part of an international transport corridor North-South tension Baltic Sea - Poland - Slovakia - Hungary - Balkans.

- A railway going through the Kechnec is included in international agreements on the European International Combined Transport Lines. This railway connects Poland, Slovakia and Hungary.
- The rail freight in Ukraine requires access to a transfer point from the standard European railway to the Eastern broader. Such a transfer point is located 3.5 km from the industrial park Kechnec [3].

4) Infrastructure:

- Electricity on the border of the park - 22 kV and 110 kV.
- 6 transformer on the border of the park (each capable of generating 1 MW).
- New energy center (80 MW).
- The gas pipe in the park with a diameter of 150 mm (high pressure) and 225 mm (medium pressure).
- Water source capable of producing 21 liters per second; pipeline with a diameter of 225 mm.
- Sewage with diameter 1000 mm.
- Capacity of sewage treatment plant is 12 000 equivalents, currently, there it is only used for about 25 to 30%.
- Industrial water - 60 l / sec; pipe with a diameter of 225 mm.
- Whole park has a stormwater sewer with sufficient capacity.
- Digital Telecommunications Center - a sufficient number of telephone subscribers and subscribers to broadband internet [3].
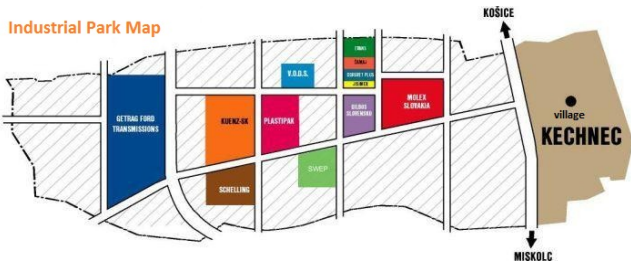
Figure 1: Map of the Kechnec Industrial Park

IV. COMPANIES ESTABLISHED IN KECHNEC INDUSTRIAL PARK

IEE Sensing Slovakia

IEE is an innovative manufacturer of intelligent scanning systems. Company with company management based in Luxembourg, was founded in 1989 and operates in Europe, USA and Asia. Scanning systems manufactured by the company are designated for: transportation and automotive industries, public and commercial infrastructure, consumer electronics, automation and logistics and medicine and healthcare. IEE is a world leader in safety sensing systems in cars to detect and evaluate the presence of a passenger, with products such as Seat Belt Reminders (SBR) and BodySenseTM.

The company began production of Seat Belt Reminders in Kechnec in 2003, working with the U.S. firm Molex. In September 2009, IEE has established its own manufacturing plant for SBR under the auspices of its Slovakian subsidiary IEE Sensing Slovakia. Since 2007, the company also implemented testing of seats and validation testing of Occupant Classification (OCTM) products in its own test lab [4].

IEE Sensing Slovakia produces annually in Kechnec about seven million SBR sensors for automobile manufacturers as Daimler, Fiat, Ford, Opel, PSA, Renault and VW. IEE Sensing Slovakia company takes the space of the company Molex Slovakia.

Molex Slovakia

The first major foreign investors in Kechnec was an American company Molex. After intensive preparations, the cornerstone of the new operation was laid on the 11th May 2000. Inauguration of the first phase of construction was about 7 months later, on 15th December 2000. The second and third phase of construction was completed in 2003.

Molex produces worldwide more than 100 000 different connector systems for computer, telecommunications and automotive industries. Molex Slovakia, OJSC Kechnec focuses on the production of connector systems for the automotive industry and is the largest manufacturing plant in the car division of Molex. Among the main customers of products produced in Molex Slovakia are Valeo, Siemens VDO, Delphi and Yazaki [3].

Getrag Ford Transmissions Slovakia, s.r.o.

Getrag Ford Transmissions raised the ceremonial stone of its new plant in Slovakia on 7th July 2005. GETRAG group, consisting Getrag and Getrag Ford Transmissions, is a specialist in drive technology and is the largest independent producer of mechanical transmissions in the world. With 12 400 employees, the company achieves an annual turnover of almost EUR 2.4 billion in its 25 manufacturing plants in Europe, Asia and USA[5].

With the products on the latest technical level and progressive manufacturing technologies, the company achieves leadership on the market of transmission systems of tomorrow. The cornerstone of its success are qualified and committed employees, both men and women.

GETRAG FORD Transimissions built in the industrial park near Kosice Kechnec technically advanced plant. In November 2006, began production of motorcycle gears began, in autumn 2007 production of dual-clutch gearbox was launched.

Currently around 760 employees are working for us, and we plan to increase the number to more than 900 in a few years. The company Getrag Ford Transmissions Slovakia sro is a subsidiary of Getrag Ford Transmissions, a joint venture (50:50) of Getrag company [2], specialist in gearboxes and Ford Motor Company.

Getrag Ford Transmissions is located in Cologne (Germany) and manufacturing facilities in Bordeaux (France), Halewood (Great Britain), in Cologne and Kechnec (Slovakia). In addition there is a development center for products in the city of Gothenburg (Sweden) [3].

V. The Use of Solar Panels to Decrease Energy Dependence of the Industrial Park

Solar energy is the cleanest source of energy we can get on earth. Based on this fact I decided to propose an alternative use of solar panels as a substitute for obtaining electricity. These panels can be installed on not built area, as well as built-up area, specifically on concrete roofs of factory buildings located in the industrial park. Such a method of saving energy on one hand, saves money, on other hand helps to the energy independence of the industrial park and last but not least is involved in preserving nature for future generations.

This way of gaining power, is (initial cost) expensive, but in long term the costs very quickly return. The guarantee of the return of money invested is essentially one hundred percent, because the use of solar panels is a global trend and this trend is supported by the State, as well as EU itself.

The buying price is currently € 0.43027 / kWh which is about 400% more expensive than the electricity from the distribution network. It is therefore necessary to realize how easy it is to earn and save at the same time. Recognise that the Act No. 309/2009 National Council comits the distribution network to buy this enrgy from you for a good price, currently € 0.43027 / kW, ie about 400% more expensive than electricity from the distribution network [1]. This offers greater government guaranteed annual returns for investors, municipalities and cities than even the best banking sector offers. The Act also provides 15-year guarantee purchase prices for all investors [6].

VI. Estimated Annual Production of Energy in the Industrial Park

The cornerstone of solar photovoltaic systems are photovoltaic fields, thus an interconnected system of photovoltaic panels and other elements such as smart power inverter (inverter), the elements of harmonic filtering, phase correction, security features, or batteries and meters of the amount of electric power.

In Slovak latitudes solar photovoltaic system or solar power plant installed with capacity of 1000 Wp produces from 950 kWh to 1150 kWh of electricity per year. Approximate price of investment in such a system is approx. € 2 500 per 1000W installed capacity, producing approximately 1100 kWh of electricity per year and income from the distribution system approx. € 470 per year. The upper limit of installed capacity practically does not exist, it is only limited by the terms of the distribution network. For example, with installed capacity of 1MW the investment is about 2.5 million € and annual revenue of about 470 000 € [7].

At current guaranteed prices 0.43027 € / kWh the return on investment is six years, after this period you get more electricity for min. 19 years virtually free, since the photovoltaic system is powered solely from its own resources. Another advantage is guaranteed security for the solar systems provided by different companies, which ranges from 20-25 years. Economic benefits of solar power plants are therefore very high, annualy 13 to 22%.

1) Solar power plant on – grid, system to supply electricity from light source to the public, or distribution network
2) Solar power plant off – grid, system to supply electricity from light source to own objects, mostlt where you can not connect to the public electricity network.

The power of panels is expressed by so called peak power (Wp). Watt is a unit used to express the ability of facilities to generate electricity, or reflects the ability of electrical equipment to consume electricity. 1 Wp is the power of the device in the intensity of solar radiation 1000 W/m^2 reaching the cell at a nominal temperature of 25 °C [4].

These conditions are obtained in good weather when the sun is at the highest point in the sky. So if we calculate, using the simple formula, the power of solar power plant, which would occupy only a hundredth of the industrial park area (approximately 33,200 hectares), annual earnings would be certainly not negligible [5].

33,200 m^2 x 800 kWh / m^2 / year = 26,560,000 kWh / year 26.56 GWh / year

This would provide an annual profit of 11,420,800 € per year. Price for direct supply of electricity to the grid is fixed and guaranteed by the state for 15 years. Therefore, it is one hundred percent guaranteed return on investment. Return on total investment is estimated to 6-7 years. Performance of 80% is guaranteed for 15 years and therefore, even if such investment is costly and time consuming its return is undeniable. Another advantage is clean energy produced, without any by-products, waste. Finally, also the profit which is within the timescale of 20 years not negligible.

VII. Conclusion

Price for direct supply of electricity to the grid is fixed and guaranteed by the state for 15 years. Therefore, it is one hundred percent guaranteed return on investment. Return on total investment is estimated to 6-7 years.

Performance of 80% is guaranteed for 15 years and therefore, even if such investment is costly and time consuming its return is undeniable. Another advantage is clean energy produced, without any by-products, waste. Finally, it is also the profit which is within the timescale of 20 years not negligible.

References

[1] D. P. Chynoweth, J. M. Owens, and R. Legrand, "Renewable methane from anaerobic digestion of biomass," *Renewable Energy*, vol. 22, no. 1-3, pp. 1 – 8, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/B6V4S-411W3CG-1/2/9b4606e1a5b7b4e943d87f8b9d13531d

[2] M. Pehnt, "Dynamic life cycle assessment (lca) of renewable energy technologies," *Renewable Energy*, vol. 31, no. 1, pp. 55 – 71, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/B6V4S-4G4XBR8-1/2/918beff8c122a8e3a7addc8da40ea738

[3] "Kechnec: The town Homepage," *Internet Source*, 2011. [Online]. Available: http://www.kechnec.sk

[4] A. Smola, *Solar Energy and its Exploitation (In Slovak)*. STU, 1994.

[5] J. P. Painuly, "Barriers to renewable energy penetration; a framework for analysis," *Renewable Energy*, vol. 24, no. 1, pp. 73 – 89, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/B6V4S-42R0S25-6/2/ef82f99b025a2a481956e5752886de88

[6] "Slovak Innovative and Energy Agency," *Internet Source*, 2011. [Online]. Available: http://www.siea.gov.sk

[7] D. Bahnemann, "Photocatalytic water treatment: Solar energy applications," *Solar Energy*, vol. 77, no. 5, pp. 445 – 459, 2004, photocatalysis. [Online]. Available: http://www.sciencedirect.com/science/article/B6V50-4D5X5S5-4/2/ea7629c1ee2e33afc851c9d966b5011d

Robot with a sensory system based on LEGO System

Marcel Vidiečan

Supervisor of bachelor work: prof. Ing. Dobroslav Kováč, CSc.

Department of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

marcel.vidiecan@student.tuke.sk

Abstract— **The aim of this work is to construct and programming model of intelligent robot sensory system based on LEGO. Represents a useful laboratory model, whether industrial or logistics for handling objects unattended, and for their further displacement in the manufacturing or shipping process. The robot s built from LEGO Mindstorms, which is designed for the construction of automated robotic models. Control is provided by the control unit NTX, which it can upload any proposed program, if necessary. Is driven by three servomotors and sees its surroundings through three types of sensors such as colour sensor, ultrasonic sensor and pressure sensor.**

Keywords— **Sensory robot control unit, sensors, actuators, software.**

I. INTRODUCTION

Factory robots and robots with sensory systems are used around two decades. Their use is largely represented mainly in the automotive industry, but over time they started to get into other spheres such as production, working materials but also manipulation of objects without operator assistance, and not least the sensory robots have become increasingly even household.

II. BACKGROUND

When drawing up the sensory robot is needed to solve many problems. Mechanical parts, which are robots made of different shapes and not all are compatible. Since the mass-produced, can be expensive or unavailable. Solve this problem LEGO Company, which in 1995 commissioned and launched LEGO technic. This step is gradually supplementing and modernizing parts of the market received a large number of compatible and affordable components. Further modernization and the need to control a robot designed by the company came with the new LEGO Mindstroms, which allows almost any model to build a robot to demonstrate under laboratory conditions. The primary part of the kit is a so-called smart cube NTX, which is within us includes a control unit with memory, allowing it to upload to a program created in a graphics program, LEGO Company which comes with a control unit via USB or Bluetooth connection to computer. Other components are modular systems such as touch sensors, light, ultrasonic. Using sensors, the robot is able to perceive their environment and obtain information about it. For the robot are designed interactive servo motors which are controllable number of turns or angles from 1 to 360 degrees. Interactive servo motors and sensors are attached to the control unit with six venous cables.

III. TECHNICAL SPECIFICATIONS

Intelligent Cube NTX is operated with 32-bit AMR7 micro control unit, 256 Kbyte Flash, 64 Kbyte RAM. Communication with PC is provided with Bluetooth Class II. through which it is possible not only to communicate with PC but also to remotely control the robot to 10 meters or USB port, 12 Mbit / s. The control unit has 4 input ports for sensors that are connected to a control unit using a 6 wire cable digital platform (One port includes IEC 61158 type 4/EN 50 170 compatible port for future use). Servomotors are used to control three output ports, 6-wire cable digital platform. The control unit has an LCD graphic display 100 x 64 pixels. Sounds used for speaker with sound quality 8 kHz. Audio channel has an 8-bit resolution and 2 to 16 kHz sample frequency. NTX cube is in itself basic software to control various sub-programs, testing and calibration programs. NTX cube can also be controlled using the buttons and navigating the menu. Power is supplied from the adapter or AA batteries or special LEGO battery
.

Fig.1 - Intelligent Cube NTX

IV. DISTRIBUTION OF SENSORS AND MOTORS

A. Sensor colour

Fig.2 Sensor of colour

Thanks to its construction, the sensor can be used for the recognition of six colours, detects the intensity of light or to measure the intensity of light coloured surfaces. The sensor may serve largely in deciding a robot. Program is possible for the robot to enter a variety of outputs, which carried by a selected colour. The sensor can also shine in four colours red blue green and white. This feature can be used to function as a robot, each colour can be programmed a certain property. For example such sensor can be used for mobile robots to determine the sensor path after which the robot to go.
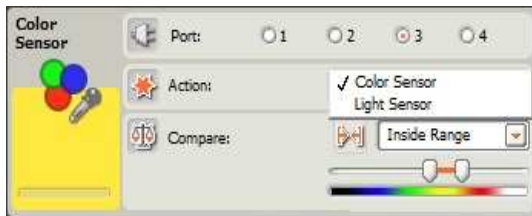
Fig.3 Palette colour setting properties of the sensor

B. Touch Sensor

Fig.4 Touch Sensor

Touch sensor provides the robot ability to touch. The sensor operates in three positions as press, release, or at once. For each property that can be programmed a certain input.

Fig.3 - A variety of setting properties of the pressure sensor

C. Ultrasonic sensor

Fig.5 Ultrasonic sensor

The ultrasonic sensor is used to detect the distance of objects to avoid obstacles, tracing objects. Sensor measures distance in centimetres or inches, which may NTX controller display to display. It can measure the distance from 0 to 255 cm with an accuracy of + / - 3 cm. The sensor uses the scientific principles of sound transmission and reflection from obstacles. Followed by calculating the period during which the reflected sound will come back to the sensor the distance is determined.
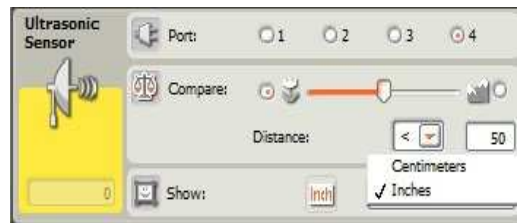
Fig.6 A variety of adjustment features an ultrasonic sensor

D. Interactive Servo Motor
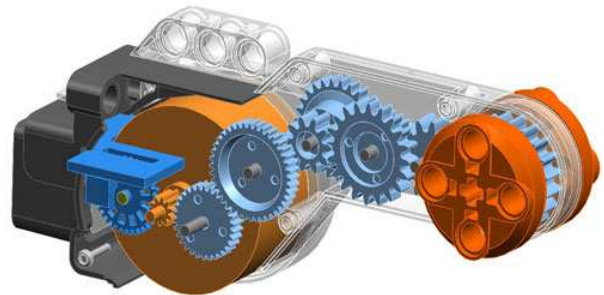
Fig.7 – Interactive Servo Motor

The robot provides three interactive servo motors. They are already constructed for the transfer of sufficient strength. Servo properties are set different performance aspects of the program on your PC. Each engine has a built-in rotation sensor that provides precise robot motions on both sides. The rotation sensor measures the rotation in degrees of engine rotation or independent on the basis of complete revolutions. One revolution of the engine corresponds to 360 degrees with an accuracy of + / - 1 degree. Actuator can be set to infinite rotation, which ends with one of the sensors or rotation time in seconds and turning in grades. The PC software is set rotational speed of servo motor and also the strength with which it works. Motors are connected to an intelligent glance at ports NTX ABC.

Fig. 8 - The range setting of Servo Motor properties

V. HOW DESIGNED ROBOT OPERATES

Sensory robot model is designed to manipulate objects in the automotive industry. Already mentioned the robot by three motors. The first actuator is connected to the control unit NTX on A output port and it controls the rotating robotic arm in the direction of right and left at an angle of 360 degrees through transfer to streamline its strength. Vertical bearing arm is attached to the bottom of the robot to transfer actuator A. The vertical arm can move the horizontal arm in the direction up and down at an angle of 150 degrees. It is operated actuator, which is connected to output port B to control unit NTX. Mass shoulders require gearing and this part. The upper horizontal arm incorporates a servo motor to open and close a robotic hand, whereby the robot can grasp objects. The servomotor is connected to the output port C in the control unit NTX.

Robotic arm can be changed as appropriate depending on what object should be transferred. The robotic arm for carrying objects in special containers uses this case. Sensors are used for robot perception. As the first, which triggers the whole process is an ultrasonic sensor. The program enables the robot when the robot approaches an object over a distance of 3 cm and ceases to move. The sensor is placed in front of the bottom of the robot. The sensor is connected to the input port 1 controller NTX. If not close to any object or robot vehicle, the robot is in stand-by mode. If the robot is close to a container, the robot program is run, causing the robot tipped over the container to fetch the container, which is. Transported containers are marked in colour. The robot is set to distinguish three types of colour and the red, green and blue. This feature ensures colour sensor that recognizes six kinds of colours and is connected to the input port 2 controllers NTX. If the container is red, the robot picks up the container and moved it to his right side and deposited him there. If the vehicle is a container with green, the robot picks up the container again and put it on the left side. Repository is used for the last type of sensors and pressure. Sensors are in the starting position extended. If the robot is positioned on the sensor storage container is inserted. This robot acquires information that the repository is busy. If container transports the vehicle, the vehicle is marked in blue and the robot receives information that the container is loaded. Loading the container is again ejected pressure sensor for robot repository and gathers information that does not contain any cost, and thus returns to the starting position.
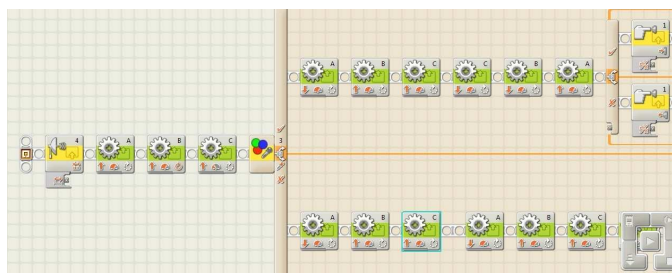
Fig.9 Example of program LEGO Mindstorms

VI. Programming

Creating programs for smart cube NTX is secured using LEGO Mindstorms NTX 2.0 software. The software is designed for easy connection to PC via Bluetooth or USB. Use the icons with the logic operation is ensured decision robot. Basic icons are icons for the movement of engines, adjusting their properties, recording sound effects, making loops and switches. Each actuator can set different properties than the length of motion, strength, speed, and port connectivity. For interactive servo motors feature is implanted base, that in case of blocking actuator program immediately assess the situation and on that basis is able to prematurely terminate the program, to avoid damaging the hardware. It is also possible to set the rotation of actuator properties. Sensors also have their icons, which are set in their properties. For the ultrasonic sensor can be set up motion detection or measuring distance. Colour sensor can be set using software as a coloured lamp that can

illuminate four colours, or as light intensity sensor recognition and six colours. For the pressure sensor can be adjusted his position, compression, impact and eject. To decide the robot and the processing of input information used e-switch or loop. If you have a robot perform a task, it is necessary to enter basic information requirements and using that to compare them with the information input. Loop is used to re-perform a specific task, until there is a change of input. The program can still use the pause, which can be adjusted by the length of time required for stopping the process or a variation of the input. The software also contains logical and mathematical comparison, randomly generated functions. The last part is the servo calibration and reset to its starting position. In compiling the program to deploy icons on the desktop and, where appropriate, shall be connected to each other or to create e-switch or for more sub-loops. Programming is based on the base of tree diagrams.

Conclusion

The main task of my work was to construct a model of sensory recovery robot technology production line for production of automobile components, using the LEGO Mindstorms.

Robot could be used in practice in logistics processes objects transported in an automated process using the minimum operator. Adding more moving parts other sensors can expand properties of the robot. Created by the requirement to use the next time you synchronize the control unit and programs running on them at the same time. This task is beyond the described problems

Acknowledgement

References

[1] LEGO Mindstorms User Guide, 2009.
[2] Web source http://dal.hnonline.sk/c1-44254690-skladove-priestory-obsadzuju-automaty
[3] Web source http://neuron.tuke.sk/~stupakp/Moje/ROBOTY.htm

Soft switching DC converter for photovoltaics

[1]Erik EÖTVÖS, [2]Tomáš BÉREŠ, [3]Marek PÁSTOR

Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]erik.eotvos@.tuke.sk, [2]tomas.beres@tuke.sk, [3]marek.pastor@tuke.sk

Abstract—This paper describes a step-up DC converter, which main task is to increase the voltage from a low level drawn from a photovoltaic (PV) panel, to a high controlled level for connected inverter. The converter also provides an electrical isolation of PV panels from the grid. In the converter a LLC topology is used. The converter operates at high switching frequency to achieve small size of the power transformer. The main benefit of this converter is zero-voltage switching (ZVS) of primary MOSFETs and zero-current switching (ZCS) of rectifier diodes over the entire operating range. Advantage of used topology is that the converter can be controlled by a simple 8bit microcontroller (MCU). A laboratory model with maximum 95.5% efficiency was built to verify properties of the LLC resonant DC/DC converter.

Keywords—PV panel, LLC resonant converter, ZCS, ZVS.

I. INTRODUCTION

In nowadays is consumption of fossil fuels on its maximum level, and new sources of oil or gas are discovered only rarely. And therefore we must thinking of how we will compensate this deficit. One of options is using renewable power sources. Today is the power of water and wind most used. However with increasing development in photovoltaics the solar energy is more and more used nowadays. This increase is related to increasing efficiency of transformation of solar energy to electrical energy. Now, the common efficiency of PV cells is over 15%, and in the laboratory was achieved efficiency up to 42%. Rising demand on a market for photovoltaic is related to reducing cost of PV panels and benefits which many countries offer. In Fig. 1 the year cumulative installed power of PV panels in years 1995 to 2008 is shown.

For using the energy drawn from PV panels, we need special type of a converter. The type of the converter depends on method how we use PV panels. If system of PV panels and converter is not connected to power grid, we talk about off-grid system. If the energy is not used, the energy obtained from PV panels is converted to charge batteries. If the consumption of energy begins, the converter starts to transfer the energy to the load through the inverter. If level of power obtained from PV panels is higher than the system can offer, the converter starts to draw the energy from batteries.

Photovoltaic systems connected to electrical grid, called on-grid, are more often used today. In this case are used special inverters, which convert the energy from PV panels directly into the grid.

It is desired to use the renewable energy sources with maximal efficiency. One of the possibilities how increase the efficiency of a PV system, is to increase the efficiency of the inverter. There are quantities of inverters for PV systems on the market. Some inverters include DC/DC step-up converter, depending on whether the inverter is connected to the string of PV panels with voltage higher than the maximum value of the grid voltage. When a thin film PV panels are used, problem with leakage current will occur. Therefore it must be used converter with transformer, which provides galvanic isolation between PV panels and grid.
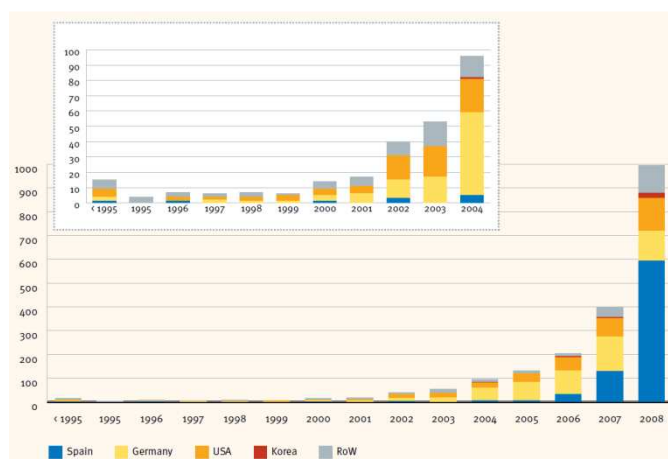
Fig. 1 Cumulative installed power of PV panels for years 1995-2008

If the line frequency transformer is used, the whole converter will be heavy and bulky. Therefore better solution can be, if transformer is used in dc converter. With high switching frequency, the transformer can achieve small dimensions. If there is second converter, the overall efficiency will decrease.
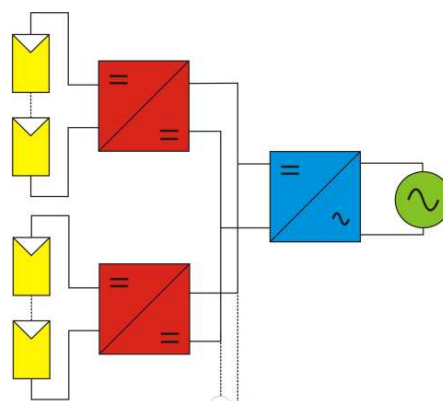
Fig. 2 Topology of converter for solar system connected to grid

But if a DC converter will be used with very high efficiency, then it is not such a problem. With this topology Fig. 2 it can be used more dc converters and there will be

achieved MPPT (Maximum Power Point Tracking) for each PV string.

The described converter should work with input voltage range from 60V to 100V. The required output voltage (the input voltage for the inverter) is 400V and the maximum output power should be about 600W.
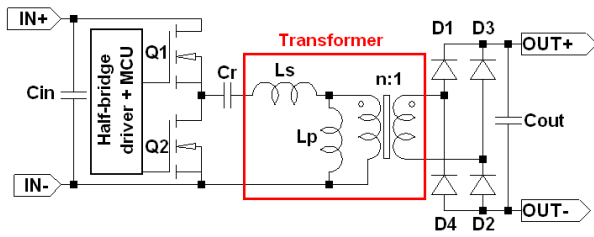
Fig. 3 The principal scheme of proposed LLC converter.

Considering that it is required to have the high efficiency and an electrical isolation the series resonant LLC half-bridge converter was chosen [1]. Principal scheme of the LLC converter is shown in Fig. 3. It consists of the half-bridge inverter, created by power MOSFET switches, from which the resonant tank is supplied.

The resonant tank of the converter consists of series "L_S" and parallel inductance "L_P" and the resonant capacitor "C_R". On secondary side of the transformer there is a full-bridge rectifier with a filter capacitor.

II. PRINCIPLE OF OPERATION

Power MOSFETs are switched with variable frequency with fixed 50% duty cycle and no overlapping. The resonant tank has two main resonant frequencies. The higher resonant frequency "f_R" depends on the series inductance and the resonant capacitor and is calculated by using equation (1). The lower resonant frequency "f_0" is additionally influenced by the parallel inductance (2). If the switching frequency is higher than "f_R", the converter operates always in the inductive area. It means that the resonant tank current lags the input voltage square waveform, and therefore switches work under ZVS condition. Below the "f_0" resonant frequency, the resonant tank behaves as a capacitive load. Therefore the resonant tank current leads the input voltage. Switches works under ZCS condition. The area between "f_0" and "f_R" is split by a borderline to the capacitive and the inductive region. The operating point in this area depends on the load of the converter. When the switching frequency is equal to the resonant frequency "f_R", the voltage gain of the resonant tank is 1. It means that converter is load independent. At normal operation condition, the operating point should be placed near to this resonant frequency. Fig. 4 shows voltage gain curves of resonant tank for few load conditions. We can see the capacitive region on the left side of the borderline and the inductive region on its right side.

$$f_R = 1/2\pi\sqrt{L_S C_R} \tag{1}$$

$$f_0 = 1/2\pi\sqrt{(L_S + L_P)C_R} \tag{2}$$

$$M = 1/\sqrt{\left(1 + \lambda - \frac{\lambda}{f_N^2}\right) + Q^2\left(f_N - \frac{1}{f_N}\right)^2} \tag{3}$$

$$Q = (\pi^2 Z_0 P_{OUT})/8n^2 V_{OUT}^2 \tag{4}$$

$$Z_0 = 2\pi f_R L_R \tag{5}$$

$$\lambda = L_S/L_P \tag{6} \qquad f_N = f_{SW}/f_{R0} \tag{7}$$
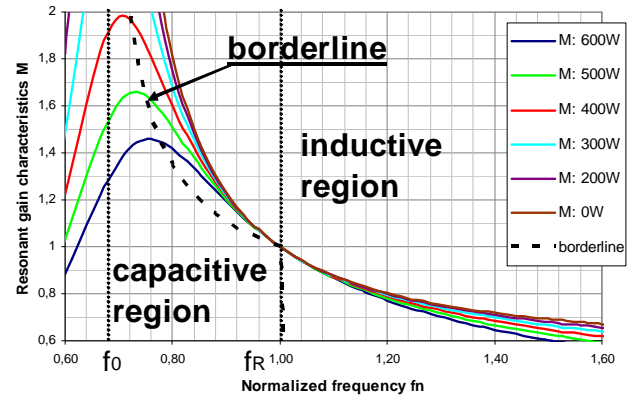
Fig. 4 Voltage gain curves of the resonant tank.

The voltage gain curves are calculated using (3), where "Q" is quality factor (4), "λ" is inductance ratio (6) and "f_N" is the normalized frequency (7).

The „Z_0"is the characteristic impedance (5) and "n" in (4) is the turn ratio of the transformer [2]-[6].

The operation of converter we can explain according to Fig. 5. in the next six phases:

1. The resonant tank current from the previous phase is flowing now through the body diode of Q1. It causes that the voltage across Q1 drops to zero, and creates the zero voltage condition for the lossless turn-on of the switch.
2. Now the current flows through Q1, and has quasi sinusoidal character. Therefore the turn-off current is much smaller.
3. Q1 and Q2 are switched-off. The current of Q1 drops to zero immediately, but the voltage across the switch rises slowly due to the charging of the output capacitance of the MOSFET. It reduces the turn-off losses.
4. Like in the first phase, the resonant tank current flows through the body diode but now of the switch Q2. The voltage of Q2 falls to zero. The switch is turned on.
5. The current flows through Q2 similarly to the second phase.
6. Switches Q1 and Q2 are switched-off again. The current of Q2 falls to zero, but the voltage rises slowly again due to charging of the transistor output capacitance [7].
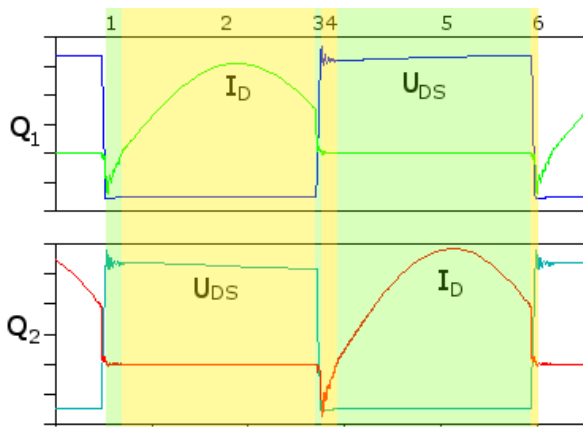
Fig. 5 Characteristic waveforms of collector – emitter voltage and collector current of half-bridge switches.

III. SIMULATION

The resonant converter was simulated in LTSpice IV program. Components of the resonant tank were calculated by equations presented in chapter II. Working frequency was set to 150 kHz. The collector current and collector-emitter voltage of the switch Q1 are shown in Fig. 6 (upper waveforms). It can be seen there that the switch Q1 starts to conduct when voltage of the switch is zero and thus ZVS is achieved for primary switches. When Q1 is turned-off, the current falls to zero, but voltages rise slowly, because the output capacitance of MOSFET is charging. On the bottom picture there are waveforms of current and voltage of the rectifier diode D1. The current through the diode starts and stops flowing when voltage is near to zero. Therefore the switching losses are minimal.
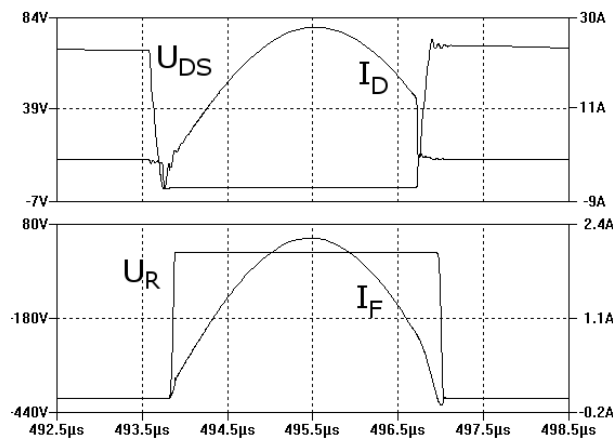
Fig. 6. Simulated waveforms of voltages and currents on switch Q1 and diode D1.

IV. LABORATORY MODEL OF THE CONVERTER

The laboratory model of the resonant converter was built. Parameters of the model:

Input voltage range V_{IN} = 60-100V.

Output voltage V_{OUT} = 400V.

Output power P_{OUT} = 600W.

When the circuit was designed, there was a need to solve few problems. Because the converter operates at high frequency and in addition with high currents, each part of circuit must by able to withstand this condition. When choosing discrete components, such as MOSFETs and rectifier diodes, we must care; that they should have minimal

loses in active mode and short switching times. But there is not problem with the maximum operating voltage, because the voltage stress is very low due to soft switching.

The input and resonant capacitors must handle very high load current at high frequency. Quality capacitors of "KPI" type are suitable to fulfil these conditions.

The transformer can be designed so that it integrates the series and the parallel inductance in one circuit. In classic transformer, the parallel inductance can by replaced by a magnetizing inductance, and series inductance by a primary leakage inductance. But there is problem with inductance ratio between magnetizing and leakage inductance, because the ratio is very small. One solution is to integrate an air gap into the magnetic circuit of the transformer. By adjusting the air gap; we can control the size of the magnetizing inductance and thereby also the inductance ratio (6). For winding we must use litz-wire to avoid a skin effect in a conductor [8].

High/Low side driver is used for driving of the MOSFETs. The driver is controlled by the 8-bit MCU. The circuit has a voltage feedback with an optocoupler.

V. EXPERIMENTAL RESULTS

The converter was connected to the DC voltage source, and loaded by an adjustable resistor. Input voltage was set to 90V. Voltages and currents of primary MOSFETs and secondary rectifier diodes were measured by a digital oscilloscope. Result is in Fig. 7. Waveforms of the voltage and the current of the MOSFET Q1 are in the top picture and waveforms of the voltage and the current of the rectifier diode are below. We can see that the current starts to flow when voltage across MOSFET is zero and thus ZVS is achieved. Moreover, the switch-off current is minimal. The current through the rectifier diode starts to flow when the voltage is near to zero. It means small switching loss. The voltage of the diode starts to rise when current falls to zero. The ZCS of the diode is achieved, too. If we look at waveforms of both voltages, we do not see any voltage spikes. It means that there is no voltage stress across MOSFETs and rectifier diodes. Therefore we can use these components with lower break down voltage, but with better other parameters.
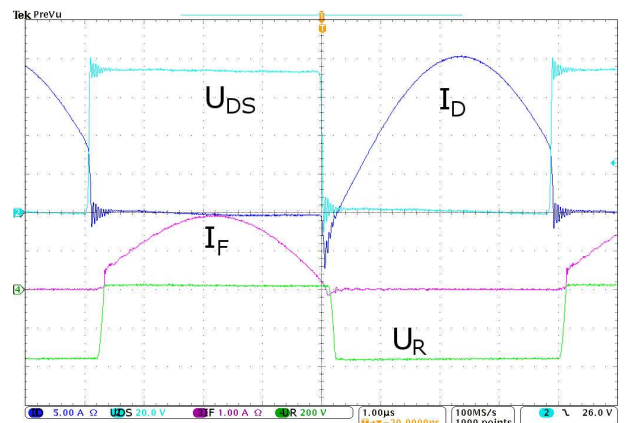
Fig. 7 Measured waveforms of currents and voltages on the primary MOSFET and the rectifier diode.

Next was measured the efficiency of the converter. The converter was loaded from 10% to 100%. Results are in a graph in Fig. 8 As we can see; the efficiency is high in wide range of load, especially over 30% load. In 50% load the

efficiency reaches its maximum, and up to 100% load, slowly declines.
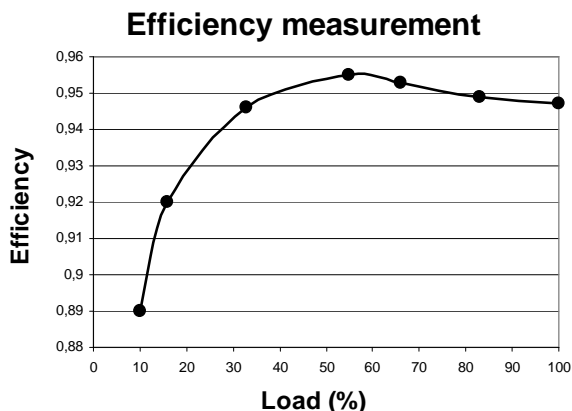
Fig. 8 Measured efficiency of converter.

VI.   CONCLUSION

The resonant converter with LLC topology has many advantages compared to other converters. Due to the high efficiency over the entire operation range, the converter is well suitable for applications such as PV systems.

ACKNOWLEDGMENT

REFERENCES

[1]   Ya Liu, "High Efficiency Optimization of LLC Resonant Converter for Wide Load Range", Blacksburg, Virginia 2007
[2]   Bo Yang, "Topology Investigation Front End DC/DC Power Conversion for Distributed System", Virginia 2003
[3]   ST Microelectronics, "LLC resonant half-bridge converter design guideline", October 2007
[4]   Mingping Mao, Dimitar Tchobanov, Dong Li, Martin Maerz, Tobias Gerber, Gerald Deboy, Leo Lorenz, "Analysis and Design of a 1MHz LLC Resonant Converter with Coreless Transformer Driver", 2007
[5]   Hang-Seok Choi/Ph.D, FPS Application group, "Half-bridge LLC Resonant Converter Design Using FSFR-series Fairchild Power Switch", 2007
[6]   Christophere Basso, "Understanding the LLC Structure in Resonant Applications", 2008
[7]   STMicroelectronics, "Simplified Analysis and Design of Series-resonant LLC Half-Bridge Converters", 2006
[8]   Fu Keung Wong, B. Eng. and M Phil., "High Frequency Transformer for Switching Mode Power Supplies", Griffith University, Brisbane, Australia, March 2004
[9]   Kácsor, G. , Špánik, P., Lokšeninec, I.: Simulation Analysis of a Zero Voltage and Zero Current Switching, DC/DC Converter. In proceedings: Transcom ´03, Žilina, Slovak Republic, 23 – 25 June 2003, pp. 47-50
[10] Hamar, J., Buti, B., Nagy,I.: Dual Channel Resonant DC-DC Converter Family. EPE Journal, Volume 17, Issue 3, Jul-Sep. 2007, pp.5-15.

14

Soft Switching DC-DC Converter

Marcel Bodor, Tomáš Béreš, Marek Pástor

Department of Electrical Engineering and Mechatronic, FEI TU of Košice, Slovak Republic

marcel.bodor@tuke.sk, tomas.beres@tuke.sk, marek.pastor@tuke.sk

Abstract—**Auxiliary circuit for DC/DC converter with controlled output rectifier is presented in this paper. Soft switching for power switches of the inverter is achieved by using controlled output rectifier and soft switching of the controlled rectifier by using lossless auxiliary circuit. The principle of converter operation is explained and analyzed and experimental results on the laboratory model are presented.**

Keywords—**ZCZVS Converter, Pulse Width Modulation (PWM), DC power supply, High frequency power converter, Snubber, Soft switching, DC-DC converter.**

I. INTRODUCTION

The conventional phase shifted PWM converters are often used in many applications because their topology permits all switching devices to operate with soft switching by using circuit parasitics such as power transformer leakage inductance and devices junction capacitance. In very used phase-shifted PWM control converters, circulating current flows through the power transformer and switching devices during freewheeling intervals. This circulating current can be eliminated by disconnection of the secondary winding, which can be realized by reverse bias application for the output diode rectifier [1] – [14] or using controlled rectifier [5], [15] – [20].

II. POWER CIRCUITS OF THE PROPOSED CONVERTER

To improve the properties of the existing converters, the new topology of the energy recovery turn-off snubber was proposed in the following DC/DC converter.

Scheme of the proposed DC/DC converter shown in Fig. 1 consists of full bridge inverter, centre tapped power transformer, controlled output rectifier, output filter and novel type of secondary turn-off snubber.

The converter is controlled by pulse-width modulation of secondary transistors. Soft switching all of the transistors in the converter is reached.

The new snubber circuit eliminates the turn off losses of the secondary transistors. The semiconductor switches T_5, T_6 in the secondary side are used to reset secondary and simultaneously also primary circulating current. The energy of the leakage inductance of the power transformer stored in snubber at secondary switch turn off is transferred to the load.

III. OPERATION PRINCIPLE

The switching diagram and operation waveforms are shown in Fig. 2 and operation analysis of the converter is shown in Fig. 3. The DC/DC converter is controlled by pulse width modulation of secondary switches.
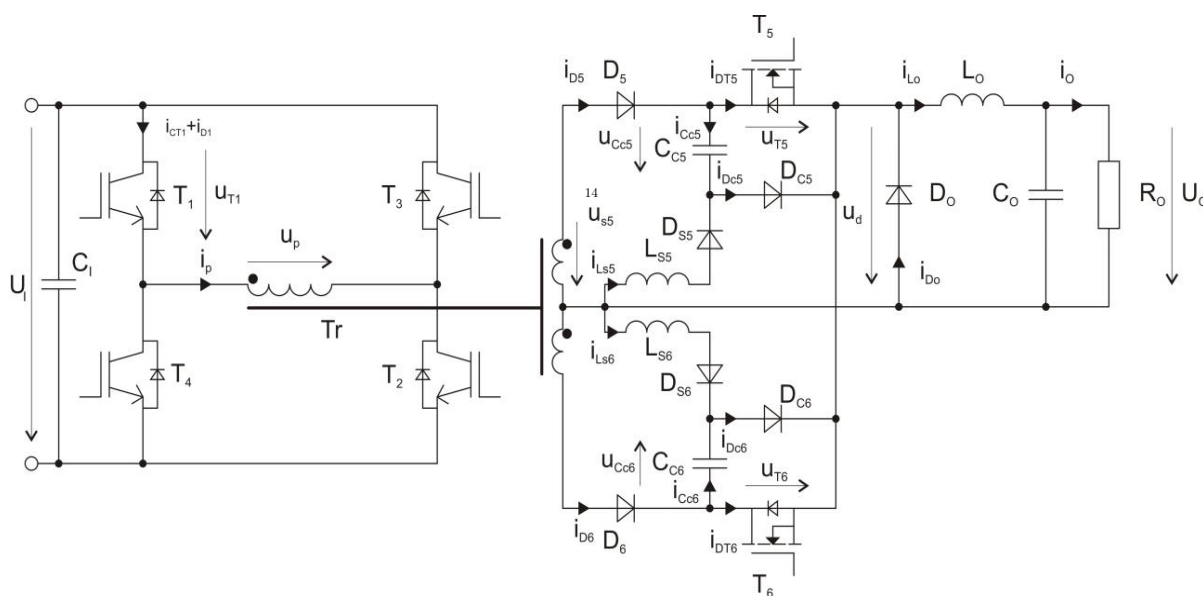
Fig. 1. Scheme of the proposed converter.
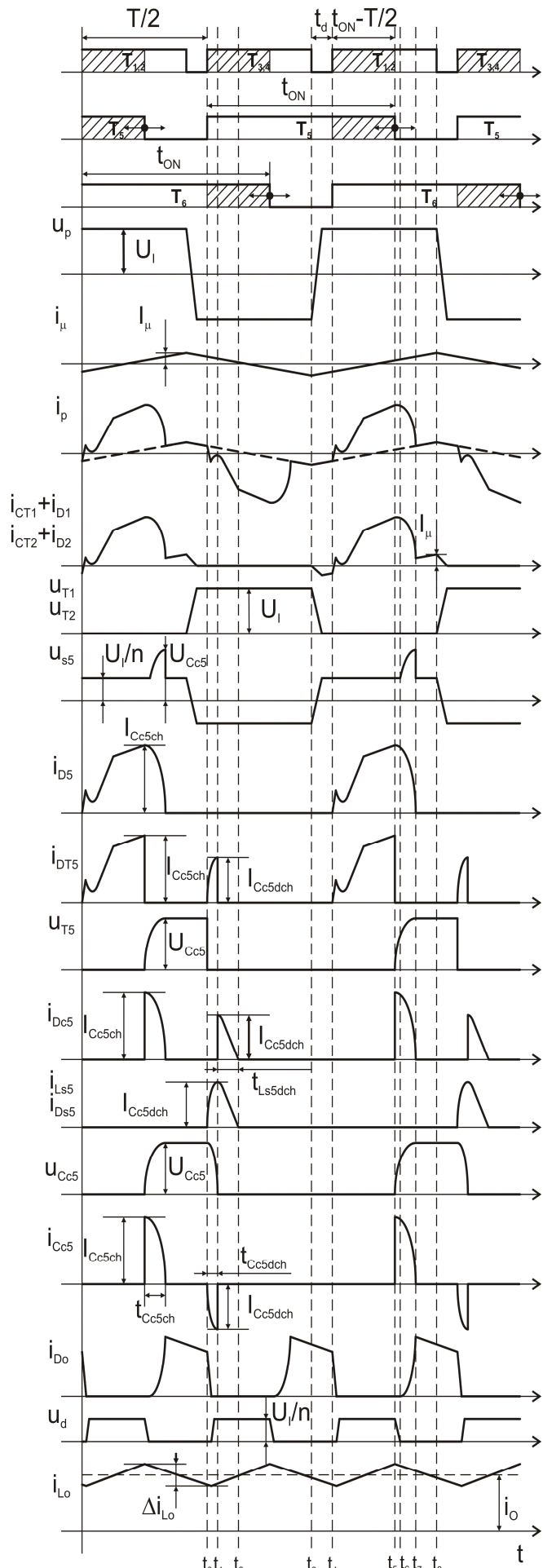
Fig. 2. Operation principle waveforms.
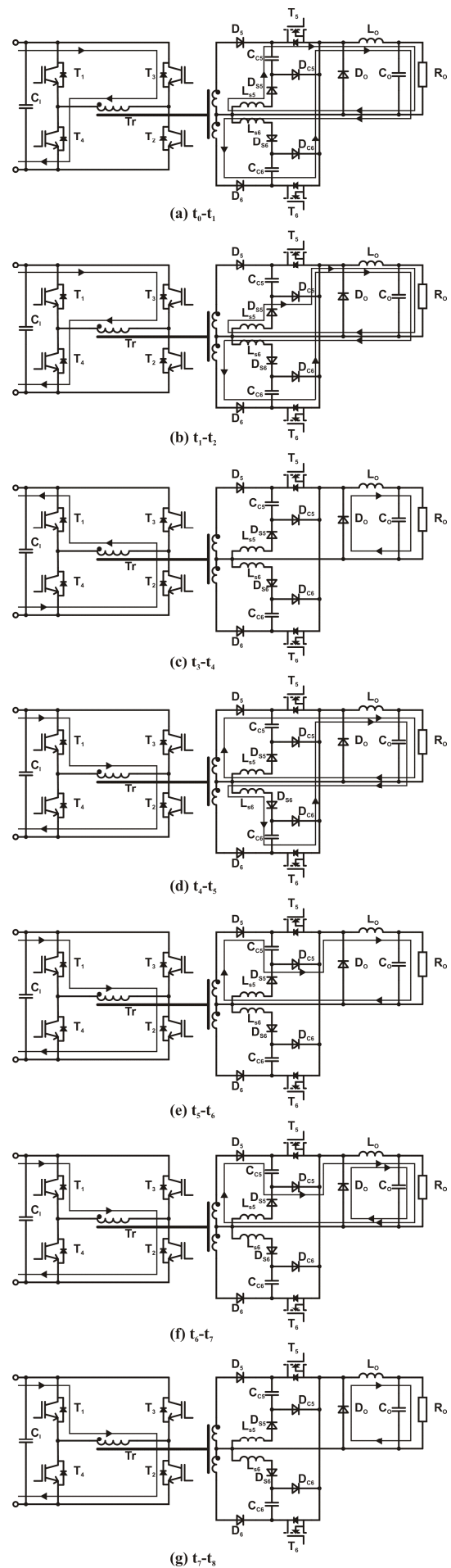
Fig. 3. Operation analysis in the intervals.

A. Interval t_0-t_1

The secondary transistor T_5 is turned on at t_0 half period earlier then primary transistors T_1 and T_2. The capacitor C_{C5} starts discharging to the load. The rate of rise of discharging current of this capacitor C_{C5} is limited by the snubber circuit inductance L_{S5}, and thus zero current turn on for the MOSFET transistor T_5 is achieved. In the same time transistors T_3, T_4 are turned on. The current of the primary transistors T_3, T_4 and the current of the secondary transistor T_6 are reduced by the discharging current of capacitor C_{C5}.

B. Interval t_1-t_2

The energy stored in snubber inductance L_{S5} is now flowing through D_{C5} to load.

C. Interval t_3-t_4

This interval starts with the turn off of the primary transistors T_3 and T_4. The magnetizing current of the transformer T_r discharges the output capacitances C_{OSS} of the transistors T_1, T_2 and charges the output capacitances of the transistors T_3, T_4. The rate of rise the current is limited by the leakage inductance of transformer. Soft turn on for transistors T_1, T_2 is achieved.

D. Interval t_4-t_5

The turn on of the transistors T_1, T_2 and T_6 commutations from freewheeling diode D_0 to T_5 transistor occur at t_4. The current of transistor T_5 is reduced by the discharge current of the capacitor C_{C6} and later by the current of the inductance L_{S6}.

E. Interval t_5-t_6

At the time t_5 transistor T_5 turns off. Its current commutates on capacitor C_{C5} and diode D_{C5}. Zero voltage turn off of this transistor is ensured because the rate of rise the voltage is limited by the capacitor. The energy of the leakage inductance of the power transformer is absorbed by the snubber capacitance and then by the load.

F. Interval t_6-t_7

At t_6 the rectified voltage u_d reached zero and afterwards the waveform of the charging process of the C_{C5} capacitance are changed. In this interval the energy of the leakage inductance is absorbed only by the capacitor C_{C5}. At t_7 the current of the rectifier diode D_5 falls to zero, and the primary current that flows through the transistor T_1 and T_2 drops on value of the magnetizing current because the whole energy was absorbed by the capacitor.

G. Interval t_7-t_8

Only the magnetizing current flows through the primary winding of the power transformer in this interval. This small magnetizing current is turned off by primary switches and thus zero current turn off is achieved. The current of the smoothing inductance L_O is flowing through the freewheeling diode.

IV. EXPERIMENTAL RESULTS

Laboratory model with components shown in Table I was built to verify the operation principle of the converter. The converter was supplied from DC source with a value of 300V. The rated output power was 1.2kW at switching frequency of 50 kHz. The typical converter waveforms were obtained at output voltage of 40V and output current of 25A.

TABLE I.
USED COMPONENTS IN CONVERTER

T_1-T_4	IRG4PSC71UD
T_5,T_6	IRFP90N20D
D_5,D_6	UFB200FA40
D_{C5}, D_{S5}, D_{C6}, D_{S6}	CSD20060D (SiC diode)
L_{S5}, L_{S6}	24µH
C_{C5}, C_{C6}	33nF
D_O	249NQ135
L_O	44µH
C_O	470µF
Tr	coaxial transformer (turn ratio 6)

Primary transistor collector-emitter voltage and collector current with gate signals of primary and secondary transistor are shown in Fig. 4. After turn off of the secondary MOSFET gate signal the transformer primary current sink to the value of magnetizing current. This small magnetizing current is later turned off by primary IGBT transistors and thus only negligible turn-off losses occur. At the turn on moment of primary IGBT transistors the primary current flows through their freewheeling diodes and rise of current is limited by the leakage inductance of transformer. This ensures zero voltage zero current turn on. Switching trajectory of primary transistor is shown in Fig. 5.
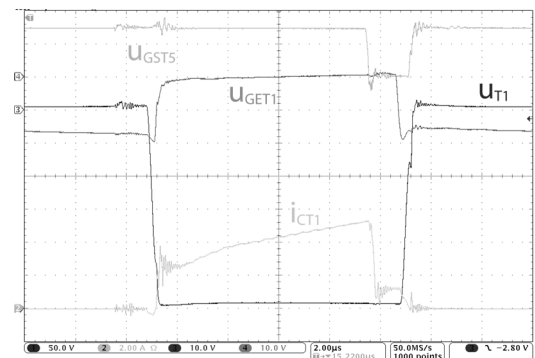
Fig. 4. Primary transistor voltage and current at turn on and turn off.
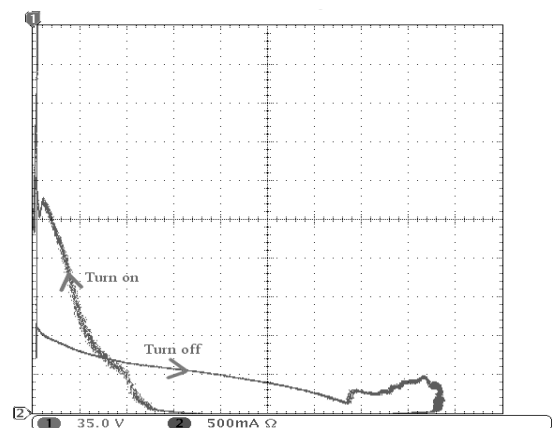
Fig. 5 Switching trajectory of the primary transistor.

Secondary transistor drain-source voltage and drain current at turn on and turn off are shown in Fig. 6. At turn off of this MOSFET transistor the collector current commutates on the snubber capacitance C_{C5} and thus the transistor voltage rate of rise is reduced. At the turn on of the transistor T_5 the capacitance C_{C5} is discharged through the transistor to the load. The rate of rise of the discharging current is reduced by inductance L_{S5}. Switching trajectory of secondary transistors is shown in Fig. 7. Charging and discharging of snubber capacitors is shown in Fig. 8.
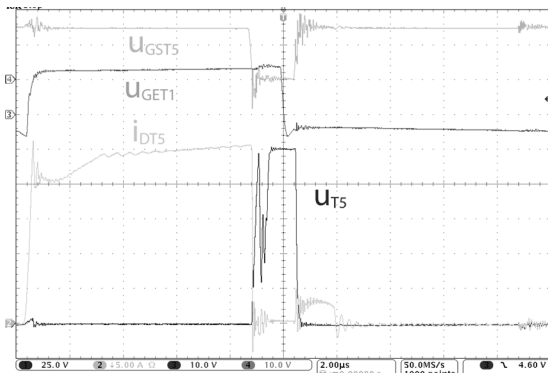
Fig. 6. Secondary transistor voltage and current at turn on and turn off.
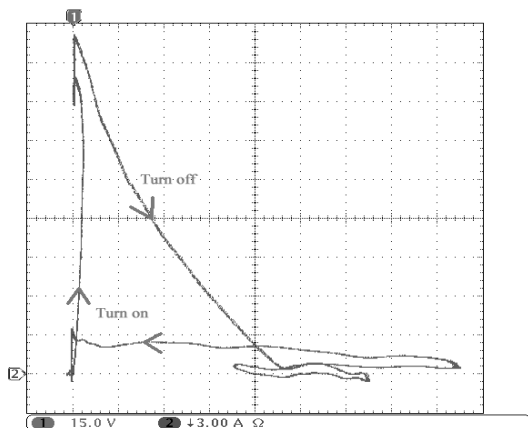
Fig. 7. Switching trajectory of the secondary transistor.
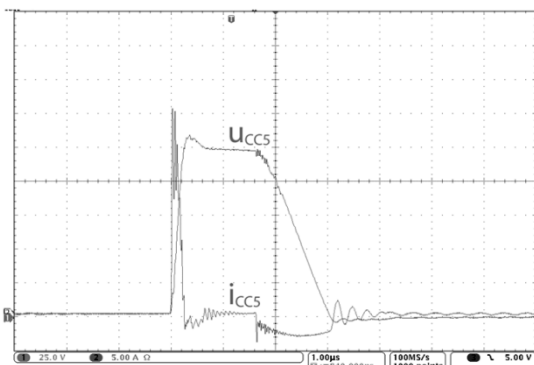
Fig. 8. Charging and discharging of snubber capacitors.

Efficiencies of the converter at various output voltages are shown in Fig. 9.
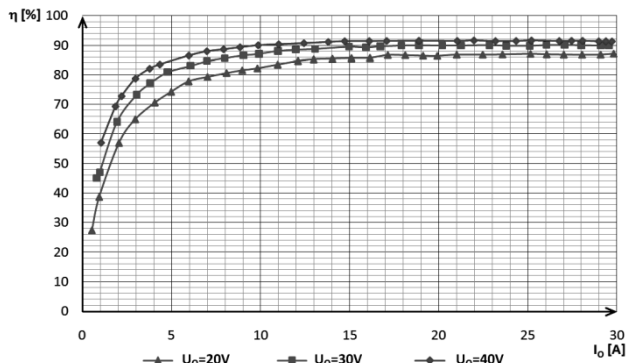
Fig. 9. Efficiencies of the converter.

V. CONCLUSION

Soft switching and reduction of circulating currents in the proposed converter are achieved for full load range using secondary side turn off snubber in combination with controlled output rectifier.

By proper design it is possible to utilize the magnetizing current of power transformer for charging or discharging output capacitances of the IGBT switches and thus zero-voltage turn-on of the IGBTs to achieve.

The IGBT transistors are turned-off almost under zero current. Only small magnetizing current of the power transformer is turned-off by IGBT transistors.

The main task of the proposed secondary turn off snubber is transfer of the leakage inductance energy to the load. Moreover it ensures zero current turn-on and zero voltage turn-off of the secondary switch.

REFERENCES

[1] J. Dudrik, P. Špánik, D. N. Trip, " Zero Voltage and Zero Current Switching Full-Bridge DC-DC Converter with Auxiliary Transformer", IEEE Trans. on Power Electronics, Vol. 21, No.5, 2006, pp. 1328 – 1335.

[2] V. Ruščin, M. Olejár, M. Lacko, and J. Dudrik, "ZVZCS DC-DC converter with controlled output rectifier", in Record, TRANSCOM 2007: 7-th European conference of young research and science workers: Proceedings: Žilina June 25-27, 2007. Žilina: University of Žilina, 2007. s. 171-174. ISBN 978-80-8070-694-4

[3] J. Dudrik, and J. Šepeľa, "Soft-switching current-mode controlled DC-DC converter with secondary switches", in Record, EDPE 2005: 13th international conference on Electrical Drives and Power Electronics, September 26-28, 2005, Dubrovnik, Croatia. Zagreb: KoREMA, 2005. 4 p.

[4] A. Tereň, I. Feňo, and P. Špánik, "DC/DC Converters with Soft (ZVS) Switching". In Record, ELEKTRO 2001, section - Electrical Engineering. Žilina 2001, Slovakia, pp. 82 – 90.

[5] J. Dudrik, "Soft switching full-bridge PWM DC/DC converter using secondary snubber". In: Journal of Electrical and Electronics Engineering. vol. 2, no. 1 (2009), p. 147-150. ISSN 1844-6035.

[6] J. Dudrik, V. Ruščin, M. Bodor, "Soft switching DC/DC converter using controlled output rectifier with secondary turn-off snubber". In: EDPE 2009: Abstracts & CD Proceedings : 15th International Conference on Electrical Drives and Power Electronics : 4th Joint Croatia-Slovakia Conference: October 12-14, 2009, Dubrovnik, Croatia. Zagreb: KoREMA, 2009. p. 1-5. ISBN 978-953-6037-56-8.

14

Spectrogram and dynamic time warping function

[1]Radoslav Bučko

[1]Dept. of Theoretical Electrotechnics and Electrical Measurement, FEI TU of Košice, Slovak Republic

[1]radoslav.bucko@tuke.sk

Abstract — **This paper describes spectrogram and dynamic time warping function as possible part of the system for recognition of voice commands for the car's equipment control such as GPS, radio, air-conditioning etc. or robotic system utilizing embedded systems.**

Keywords — **spectrogram, embedded system, dynamic time warping.**

I. INTRODUCTION

In present time the recognition of spoken speech is highly developed. Most research is focused on using the English language for such communication, however our research is aimed for using the Slovak language.

Research of recognition of spoken speech on our department is oriented on recognition of simple instructions by spectrogram. These instructions are used for car's equipment control such as GPS, radio, air-conditioning or robotic system with embedded systems.

Some problems with recognition of spoken speech are:
   a) speaker's voice can be different in various conditions,
   b) different speakers have different voices,
   c) changing environment causes trouble for speech,
   d) recorded voice can be degraded by quality of microphone or by distance from it. [1]

II. EMBEDDED SYSTEM

An embedded system is a computer system designed to perform one or more dedicated functions often with real-time computing constraints. One or more main processing cores control embedded systems. They are typically represented either by microcontrollers or digital signal processors (DSP). The program instructions written for embedded systems run with limited computer hardware resources: little memory and operating output. Because of this, it is important to optimize the acoustic signal processing used by embedded systems for speech recognition. [2]

III. TEACH-ROBOT

For control with simple instruction are selected Teach-Robot. Mechatronic system Teach-Robot is angular arm with 5 axles and 6 DC-motors (Table 1). Teach-Box provides manual control of Teach-Robot and provides communication between Teach-Robot and personal computer (Fig.1). [3]

IV. SIGNAL PROCESSING

The basic principle of most methods for acoustic signal
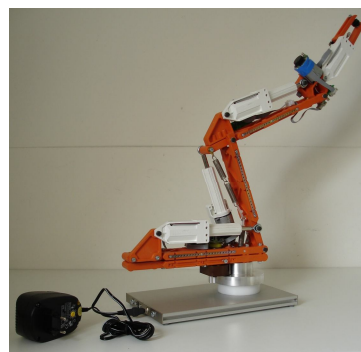
Fig.1 Mechatronic system Teach-Robot

TABLE 1
Moving specifications

Meaning	Motor	Number of pulses	Angle
grip of gripper	M1	70	60°
rotation of gripper	M2	130	200°
wrist up/down	M3	420	90°
rotation of upper arm	M4	420	90°
rotation of lower arm	M5	350	80°
rotation of body	M6	700	320°

processing is the assumption that its properties are changing slowly. Methods called short-term analysis separated and processed segments of speech signal. These segment are micro segments which are represented by the time segment of 10 to 30 ms(our system 10 ms and overlap 5 ms) . Because these micro segments are connected or can overlap each other we will get the sequence of numbers, which describes the speech.

Speech signal is recorded mostly by microphone, so the analogue signal is recorded. Analogue cycles are digitalized, that the continuous signal is represented by sequence of numbers. This process is called pulse code modulation. Pulse code modulation consist of two operations:

   - sampling in time,
   - quantization[4].

A. Processing by time

Most methods of short term analysis in time can be described by following equation:

$$Q_n = \sum_{k=-\infty}^{\infty} \tau(s(k)) w(n-k), \qquad (2)$$

where Q_n is the short time characteristics, $s(k)$ is the sample of acoustic signal get by pulse code modulation in time k, $\tau(s(k))$ is the transformation function a $w(n)$ is the weight sequence (or window) which chose the samples $s(k)$.
Hamming´s windows are used when processing in time.

Hamming´s window is defined as (fig.2)[5]:

- $w(n)=0,54-0,46cos[2\pi n/(N-1)]$
  for $0 \leq n \leq N-1$,
- $w(n)=0$ for other n.
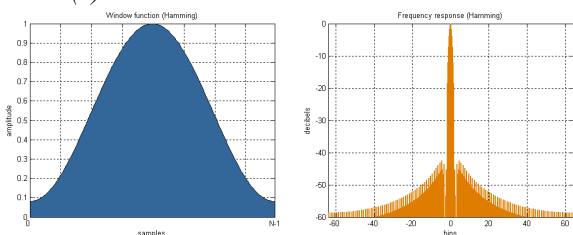
Fig.2 Hamming window

B. Short-time energy

Big problem with speech recognition is how to determine when the instruction is spoken. For this, we will apply the function of short-time energy. Function of short-time energy can be described by the following equation:

$$E_n = \sum_{k=-\infty}^{\infty} \left[s(k)\, w(n-k) \right]^2, \qquad (3)$$

where $s(k)$ is the sample of acoustic signal get by pulse code modulation in time k and $w(n)$ is in our system Hamming's windows with length of micro segment 10 ms and sampling rate 8 kHz.[6]

V. SPECTROGRAM

We have chosen spectrogram for our method of speech recognition with embedded systems. A spectrogram is a time-varying spectral representation (forming an image) that shows how the spectral density of a signal varies with time.

Figures (Fig. 3, Fig. 4) are showing the same word (word
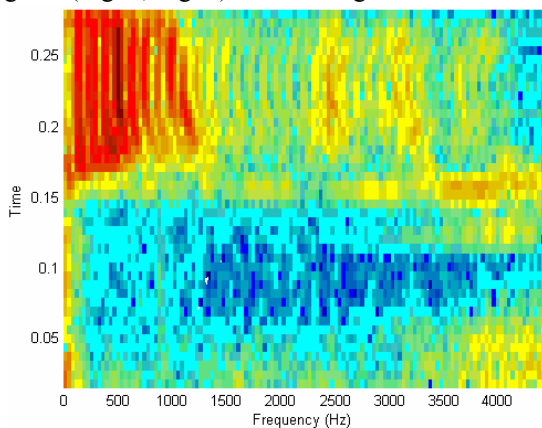
Fig.3 Spectrogram of word "stop" spoken by first person
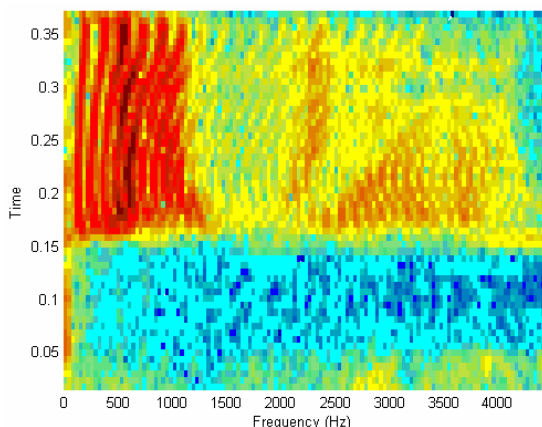
Fig.4 Spectrogram of word "stop" spoken by second person

"stop") spoken by two people. We can see similarities there. Fig. 5 is showing different word (word "vpravo"). We can see that the spectrographic picture is clearly different, which is very important for our recognition system [7].
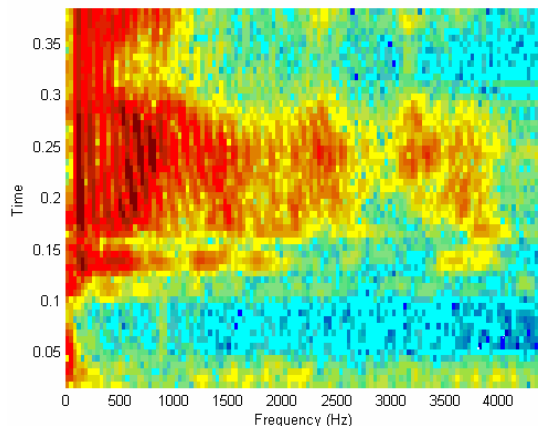
Fig.5 Spectrogram of word "vpravo" spoken by second person

We have divided the spectrogram to several sectors, in order to ease the computing process. The final value is made by arithmetic mean of these sectors (Fig.6).
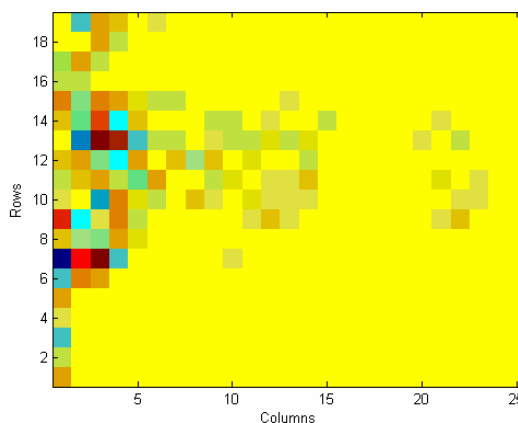
Fig.6 Spectrogram of word "vpravo" with arithmetic mean of sectors spoken by second person

VI. DYNAMIC TIME WARPING

The correctness of spoken speech classification is influenced by inconsistency and non-linear change of acoustic signal in time axis. The same word spoken by one person can have different overall length, even different length of some parts of the words (phonemes, vocals). Using algorithm of speech images comparison, which is based upon dynamic programming, can solve this problem. This algorithm utilizes non-linear time normalization. The change in time axis is modeled by non-linear dynamic time warping function (DTW) with strictly defined properties. Time differences between two speech images A and B are eliminated by deformation of one time axis, in order to achieve maximum correspondence with second image.

The image of the tested word is marked as A:

$$A=\{a(1),\ a(2),...,\ a(n),...,\ a(I)\}, \qquad (4)$$

where a(n) is n-th vector of tested word's attribute and image of reference word B is described as:

$$B=\{b(1),\ b(2),...,\ b(m),...,\ b(J)\}, \qquad (5)$$

where b(m) is m-th vector of reference's word attribute.

Algorithm with DTW function then searches for optimal path in plane *(n,m)*:

$$m = \psi(n), \tag{6}$$

which minimizes the D function of total distance between images *A* and *B* defined as:

$$D(A,B) = \sum_{n=1}^{I} \hat{d}\left[a(n), b(\psi(n))\right], \tag{7}$$

where $\hat{d}\left[a(n), b(\psi(n))\right]$ is the local distance between *n*-th vector of tested word's attribute and *m*-th vector of reference's word attribute.

We have to define the time variable *k* for which the following statements are valid:

$$n=i(k), \ k=1,...,K, \ m=j(k), \ k=1,...,K,$$

where *K* is the length of time axis for *A* and *B* pictures comparison. The situation is demonstrated on Fig.7. [1]
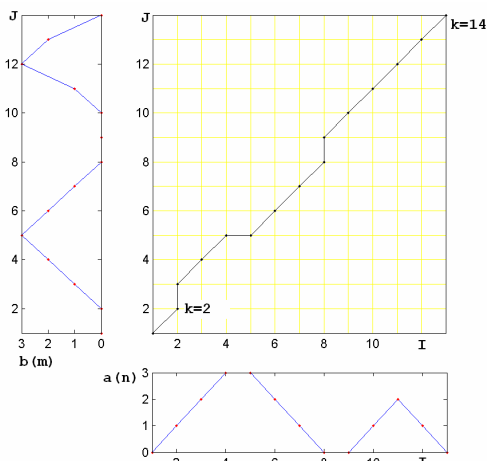
Fig.7 Comparison of tested *a(n)* and reference *b(m)* image in plane *(n,m)*

In order to find the optimal path it is essential to fulfill following conditions:

A) Limitation to start and end border points
   $$i(1)=1, \quad j(1)=1, \ i(K)=I, \quad j(K)=J.$$

B) Conditions of monotony and coherence
   $$0 \le i(k)-i(k-1) \le I^{*}, \ 0 \le j(k)-j(k-1) \le J^{*}.$$
   $$I^{*}, J^{*}=1,2,3$$

C) Limitation of DTW function steepness
   If the point *[i(k),j(k)]* moved with increasing *k* in direction of one axis, for example *n** times, the movement in this direction is forbidden unless it moves *m** times in different direction.

D) Global limitation of DTW function movement area

In order to improve the computational cost and optimize the DTW sensitivity similarly to the step function constraints global constraints were introduced. The „Sakoe-Chiba band" and „Itakura parallelogram", both global path constraints shown at Figure 8 and define the set of points available for DTW alignment only from the non-shaded regions [2].
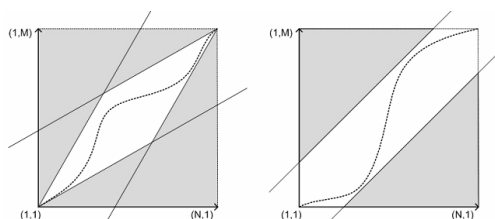
Figure 8: Global path constraints, left: Itakura parallelogram, right: Sakoe-Chiba band.

A. Weighting

By adding the weights to the each of the distances based on the step direction we could penalize or favor certain types of point-to point correspondence. 2 most types weight function (Fig.9) [3]:

- symmetric weight function:
  $$\hat{W}(k) = \left[i(k)-i(k-1)\right] + \left[j(k)-j(k-1)\right], \tag{8}$$
- asymmetric weight function:
  $$\hat{W}(k) = i(k)-i(k-1), \tag{9}$$
  $$\hat{W}(k) = j(k)-j(k-1), \tag{10}$$
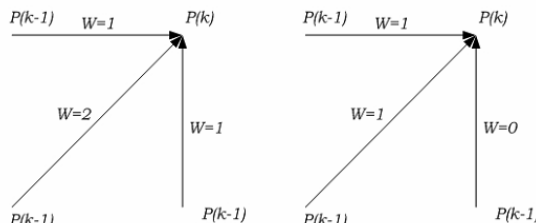
Figure 9: Weighting coefficients w(k) for symmetric (left) and asymmetric (right) forms.

VII. CONCLUSION

Spectrograms together with dynamic programming offer interesting way of speech recognition with using only a limited hardware resource of embedded systems.

REFERENCES

[1] J. Psutka, "Comomunication with PC using spoken speech" ACADEMIA, Prague, 1995.
[2] Embedded system, http://en.wikipedia.org/wiki/ Embedded_system, 18.7.2010.
[3] Teach-Robot Technical Specifications,http://www.teach-robot.nl/teachrobot2/Engels/technical_specifications.htm, 28.11.2009
[4] Čuchran, J.: "Digital transmission system". STU, Bratislava, 2008
[5] Kůs, V.: Influence of semiconductor converters on power system (in Czech). BEN technical literature, Prague 2002, 184 pages, ISBN 80-7300-062-8B.I.
[6] J. Psutka, "Speaking Czech with computer" ACADEMIA, Prague, 2006.
[7] Homemade Speech Recognition with NETcf, http://www.mperfect.net/noreco/, 15.2.2010
[8] P. Senin, "Dynamic Time Warping Algorithm Review", http://129.173.35.31/~pf/Linguistique/Treillis/ReviewDTW.pdf 10.11.

Stability of Power System (May 2011)

[1]Vladimír KRIŠTOF, [2]Stanislav KUŠNÍR, [3]Matúš KATIN

[1]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
[2] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
[3] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]vladimir.kristof@tuke.sk, [2]stanislav.kusnir@tuke.sk, [3]matus.katin@tuke.sk

Abstract— **Due to increase of electric power demands, power systems are larger and more complicated, and the dependence of people on electricity increases. Outages in electric supply have an increasing social and economic impact. Therefore it is necessary to minimize the effect of disturbances that arise in the power system so as to have minimum impact on the reliable and safe supply of electricity. Stability of power system is one of the most important area of power system operation. Loss of stability (loss of synchronism) can lead to outages of transmission lines, loss of loads, cascading failures, and eventually to black-out. This paper discusses power-system instability and the importance of fast fault-clearing performance to aid in reliable production of power.**

Keywords—**power system, steady-state stability, transient stability, fault-clearing time**

I. INTRODUCTION

Power system stability has been recognized as an important problem for secure system operation since the 1920s [3], [4]. Many major blackouts caused by power system instability have illustrated the importance of this phenomenon [3].

The term stability of the power system is connected with transient phenomena associated with changes in generator rotor angle, changes in frequency and voltage. Given the wide range of issues there is need for classification of power system stability according to Fig.3. [5]

This article deals mainly with steady-state and transient stability.

II. BASIS FOR STEADY-STATE STABILITY

In an interconnected power system, the rotors of each synchronous machine in the system rotate at the same average electrical speed. The power delivered by the generator to the power system is equal to the mechanical power applied by the prime mover, neglecting losses. During steady-state operation, the electrical power out balances the mechanical power in. The mechanical power input to the shaft from the prime mover is the product of torque and speed $(P_M = T_M\,\omega)$. The mechanical torque is in the direction of rotation. An electrical torque is applied to the shaft by the generator and is in a direction that is opposite of the rotation, as shown in Fig. 1.

When the system is disturbed due to a fault or when the load is changed quickly, the electrical power out of the machine changes. The electrical power out of the machine can change rapidly, but the mechanical power into the machine is relatively slow to change. Because of this difference in speed of response, there exists a temporary difference in the balance of power. This power unbalance causes a difference in torque applied to the shaft, which causes it to accelerate or decelerate, depending on the direction of the unbalance. As the rotor changes speed, the relative rotor angle changes. Fig. 2 shows the relationship between the rotor (torque) angle δ, the stator magnetomotive force (MMF) $F1$, and the rotor MMF $F2$. The torque angle δ is the angle between the rotor MMF $F2$ and the resultant of the vector addition of the rotor and stator MMFs R, as shown in Fig. 2
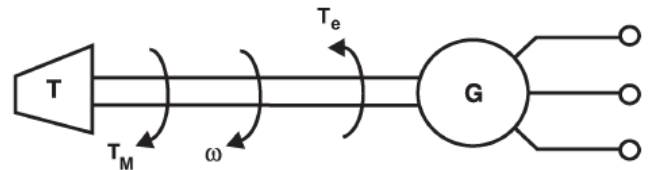
where P_M = Mechanical power
ω = speed
T_e = Electrical torque

Fig. 1. Mechanical and electrical torques applied to the shaft [2].
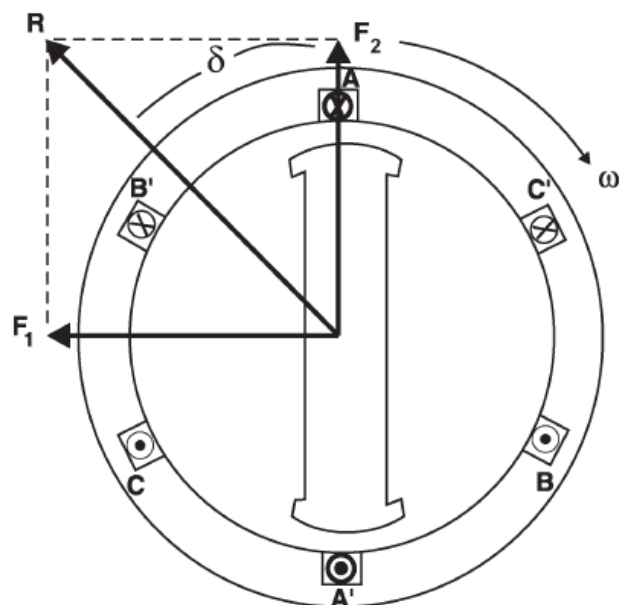
Fig. 2. Stator, rotor, and resultant MMFs and torque angle [2].
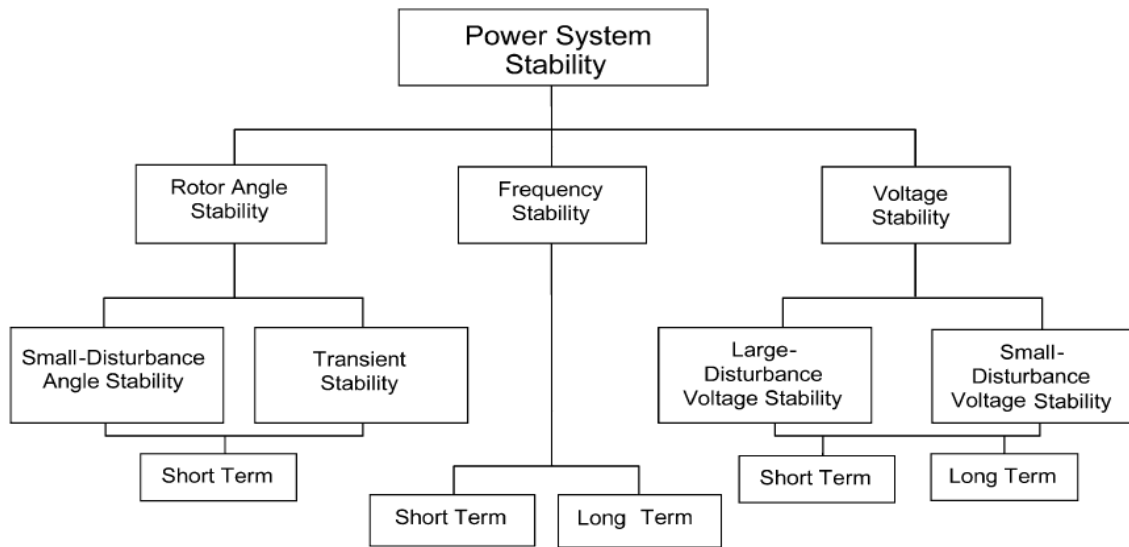
Fig. 3. Classification of power system stability according to IEEE and CIGRE [3,7]

Fig. 4 shows a circuit representation of a synchronous generator connected through a transmission system to an infinite bus. The synchronous machine is modeled by an ideal voltage source *Eg* in series with an impedance *Xg*. The terminal voltage of the machine *ET* is increased to transmission system levels through a generator step-up (GSU) transformer, which is represented by an impedance *XT* . The high voltage side of the GSU is connected to the infinite bus via a transmission line represented by reactance *XL*. The real (MW) power output from the generator on a steady-state basis is governed:

$$P_e = \frac{E_g \cdot E_T}{X_g} \cdot sin\,\delta \qquad (1)$$

by where δ is the angle between the generator terminal voltage and the internal voltage of the machine. As the power transfer increases, the angle δ increases. A fault in the system can result in a change in electrical power flow, resulting in a change in the power angle δ. This is shown in Fig. 5.

where E_T = Generator terminal voltage
Xg = Generator internal reactance
E_g = Internal generator voltage
E_{HV} = High side voltage of generator step transformer
E_O = System voltage
X_E = External impedance

Fig. 4. Synchronous machine tied to infinite bus

Fig. 5. Phasor diagram, generator tied to infinite bus

If a fault causes the current *I* to increase and the terminal voltage to decrease, the electrical power out of the machine will decrease since the impedance seen by the generator is now mainly inductive. This disturbance causes the rotor angle to increase, perhaps beyond the limits of generator synchronous operation. The resulting variations in power flow as the rotor accelerates will cause a well-designed loss of synchronism protective relaying (78 function) to isolate that generator from the rest of the system. The disturbance on the remaining system, which is due to the loss of generation, may result in additional units tripping offline and, potentially, a cascading outage.

III.   TRANSIENT STABILITY

Generators are connected to each other by a network that behaves much like weights interconnected by rubber bands (see Fig. 6). The weights represent the rotating inertia of the turbine generators, and the rubber bands are analogous to the

inductance of the transmission lines. By pulling on a weight and letting go, an oscillation is set up with several of the weights that are interconnected by the rubber bands. The result of disturbing just one weight will result in all the weights oscillating. Eventually, the system will come to rest, which is based on its damping characteristics. The frequency of oscillation depends on the mass of the weights and the springiness of the rubber bands. Likewise, a transient disturbance to the generator/network can be expected to cause some oscillations due to the inability of the mechanical torque to instantaneously balance out the transient variation in electrical torque.
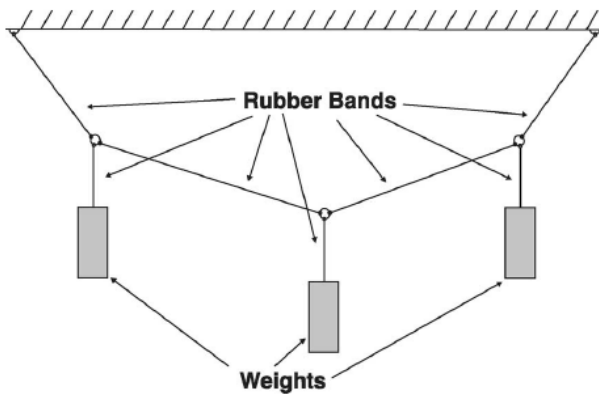
Fig. 6. Rubber band analogy

The synchronous machine's electrical power output can be resolved into an electrical torque Te multiplied by the speed ω. Following a disturbance, the change in electrical torque can further be resolved into two components:

$$\Delta T_e = K_S \Delta\delta + K_D \Delta\omega \qquad (2)$$

where:

$K_S\Delta\delta$ - component of torque that is in phase with the rotor angle change. This is known as the "synchronizing torque;"

$K_D\Delta\omega$ - component of torque that is in phase with the speed change. This is known as the "damping torque."

Both components of torque act on each generator in the system. A lack of sufficient synchronizing torque will result in loss of synchronism. Such a loss of synchronism can only be prevented if sufficient magnetic flux can be developed when a transient change in electrical torque occurs. This is facilitated by a high initial response excitation system (an excitation system that will cause a change from the input to output within 1 s) that has a sufficient field-forcing capability and a sufficiently fast response to resist the accelerating or decelerating rotor. In order to be effective for both accelerating and decelerating rotor responses, the excitation system must be capable of fieldforcing positively and negatively, particularly on generators with rotating exciters. When the rotor is accelerating with respect to the stator flux, the rotor angle is increasing due to a mechanical torque that is

higher than an electrical torque.
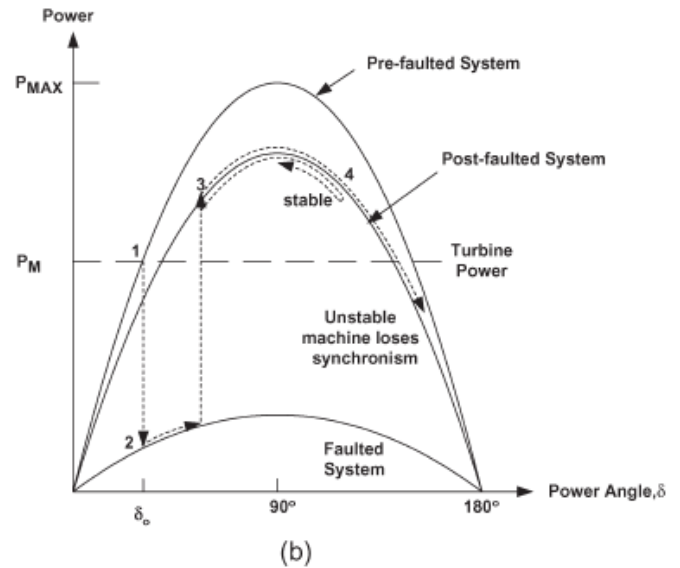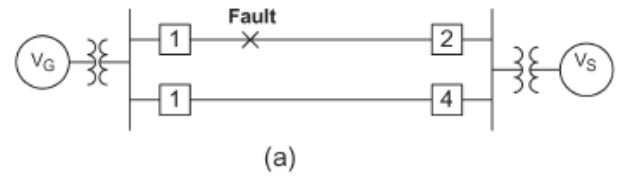
Fig. 7. Transient stability illustration: a) System b) rotor angle curve

The exciter system must increase the excitation by applying a high positive voltage to the alternator field as quickly as possible. Conversely, when the rotor angle is decreasing due to a mechanical torque that is less than an electrical torque, the exciter system must decrease the excitation by applying a high negative voltage to the alternator field as quickly as possible.

Starting from the initial operating condition (point 1), a close-in transmission fault causes the generator's electrical output power Pe to be drastically reduced. The resultant difference between the electrical power and the mechanical turbine power causes the generator rotor to accelerate with respect to the system, increasing the power angle (point 2). When the fault is cleared, the electrical power is restored to a level corresponding to the appropriate point on the power-angle curve (point 3). Clearing the fault necessarily removes one or more transmission elements from service and at least temporarily weakens the transmission system. After clearing the fault, the electrical power out of the generator becomes greater than the turbine power. This causes the unit to decelerate (point 4), reducing the momentum that the rotor gained during the fault. If there is enough retarding torque after fault clearing to make up for the acceleration during the fault, the generator will be transiently stable on the first swing and will move back toward its operating point. If the retarding torque is insufficient, the power angle will continue to increase until synchronism with the power system is lost. [2]

Fig. 8. Influence of fault clearing time: a) slow fault clearing time b) fast fault clearing time

Power-system stability depends on the clearing time for a fault on the transmission system. Comparing the two examples in Fig. 8 shows this point. In the example of slower fault clearing (a), the time duration of the fault allows the rotor to accelerate so far along the curve of PE that the decelerating torque comes right to the limit of maintaining the rotor in synchronism. The shorter fault-clearing time (b) stops the acceleration of the rotor much sooner, assuring that sufficient synchronizing torque is available to recover with a large safety margin. This effect is the demand placed on protection engineers to install the fastest available relaying equipment to protect the transmission system.

IV. CONCLUSION

Due to the increasing electricity consumption and economic and time consuming construction of new power lines, existing networks are operated more at the limits of their possibilities and they are just an ultimate of stability of power system. Therefore it is necessary pay close attention to power system stability and control.

REFERENCES

[1] Krištof, V.: *Riadenie ES v krízových situáciach*. Minimová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2010.
[2] Basler, M.J. – Schaefer, R.C.: *Understanding Power system stability*. In:Industry applications, IEEE Transactions on. March 2008. ISSN: 0093-9994
[3] Kundur, P. – Paserba, J.- Ajjarapu, V. – Andersson, G.- Bose, A. – Canizares, C. – Hatziargyrou, N. – Hill, D.- Stankovic, A.- Taylor, C. – Van Custem, T. – Vittal, V. : *Definiton and clasification of power system stability*. In: Power Systems, IEEE Transactions on. August 2004. ISSN: 0885-8950
[4] Krištof, V. – Kušnír, S.- Hlubeň,D.- Kolcun, M.: *Modeling of transient phenomena in PSLF*. In: Electrical power engineering 2010 : international scientific and technical conference : 14-16.10.2010, Varna, Bulgaria. - Varna : TU, 2010 P. 186-190. - ISBN 978-954-20-0497-4
[5] Mešter, M.: *Analýza hraničných podmienok stability elektrizačnej sústavy pri rôznych typoch skratov*. Dizertačná práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2004.
[6] CIGRE: *Advanced Angle stability controls*. CIGRÉ Technical Brochure. International Conference on Large High Voltage Electric Systems 1999.
[7] Mešter, M. – Hvizdoš, M. – Rusnák, J. – Szathmáry, P. – Vargončík, M: *Stabilita elektrizačnej sústavy*. Equilibria 2006. ISBN 80–969224– 9-1
[8] Prabha, K.: *Power System Stability and Control*. United States of America. 1993. 0-07-035958-X
[9] Machowski, J. – Bialek, J. W. – Bumby, R. J.: *Power system dynamics and stability*. England, 1997. 484 s, 978-0-471-95643-3
[10] Yorino, N. – Priyadi, A. – Kakui, H. – Takeshita, M. : *A new method obtaining critical clearing time for transient stability*. In: Power Systems, IEEE transactions on. August 2010. ISSN 0885-8950
.

Stencil Materials and Technologies for Solder Paste Deposition

[1]Michal KRAVČÍK, [2]Igor VEHEC

[1]Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic
[2]Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic

[1]michal.kravcik@tuke.sk, [2]i.vehec@tuke.sk

Abstract— **Solder pastes are key materials in surface mount technology (SMT) for assembly of printed circuit boards (PCBs). As the trend towards miniaturization of electronic products continues, there is an increasing demand for better understanding of the flow and deformation that is, the rheological behavior of solder paste formulations. Wall slip effect that is connected with requirement of very accurate (shape and volume) solder paste deposition plays an important role in characterizing the flow behavior of solder paste materials. The aim of this study is to investigate the influence of the solder paste formulation on wall-slip formation and its effect on the printability of these pastes material. Paper refers to dynamics of printing parameters for one of the most common method of solder paste deposition-stencil printing.**

Keywords— **stencil printing, solder paste, viscosity of solder paste.**

I. INTRODUCTION

Over the last 30 years, the SMT assembly process has become increasingly more sophisticated. There are two primary methods of applying solder paste to a PCB: Stencil printing and dispenzing. Which method will be used for deposition depend on the size and the type of assembled components, availability of technology, massivity of production and etc. While each method has its advantages and disadvantages, this study focuses on dynamics and different aspects of dispenser deposition on the solder paste print deposition quality. There is a wide opinion in the industry that the paste printing process accounts for the majority of assembly defects. Experience with this process has shown that typically over 60% of all soldering defects are due to the problems associated with deposition process of solder paste. Those defects faults are a major source of board failure. Unfortunately, understanding the causes of those defects can be challenging because there are numerous factors to consider. There is several factors influence to printing quality of solder paste, which has equivalent relevance [1]:

- **Environment:** temperature of solder paste during storage and during printing, room temperature, humidity, solder paste storage time, solder paste shelf time, etc.
- **Solder paste formulation and material:** alloy type, powder size, manufacture date, water soluble, viscosity, thixotropy, relaxation time, flow curve, metal content, flux type, tack time and force, etc.

- **Equipment:** type of stencil printer, technology of stencil manufacturing, printer alignment, optical inspection system.
- **Human factor:** employee skills and knowledge, part of day, operator errors and training frequency.

Stencil printing process is very alike screen printing. Prepared PCB is guyed to frame holder and cover by (mostly metal) stencil. Stencil design has apertures only at places where solder paste should be deposited. Squeegee blade moves cross whole stencil and push solder paste via apertures, like is it shown at Fig. 1.
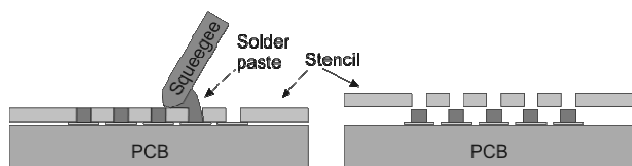
Fig. 1. Solder paste stencil printing process [1].

Like other solder deposition methods (screen printing, dispensing) also stencil printing require specific values of solder paste viscosity. Typical viscosity of lead-free solder paste for stencil printing is between 150 and 250 Pa.s [1], [2], [3]. Viscosity of solder paste is regarded as inappropriate, when:

- solder paste is sticky to squeegee blade,
- flux segregate from metal alloy,
- surface of solder paste has dry look,
- shape breakdown of deposited solder paste,
- incoming to clogging of stencil apertures by solder paste.

Advancements of this method are good repeatable, long stencil life, high speed, high accuracy and low price of solder paste deposition in big series. Stencil printing disadvantaged are relative high price of stencil, which depend on stencil material, stencil thickness and manufacture methods, and necessary stencil replace when motive is changed, and also illusoriness to deposit different height of paste. Solder paste deposition quality mostly depend on stencil quality, right rheological properties of solder paste and dynamics of squeegee setup.

II. STENCIL MATERIAL AND TECHNOLOGY

Stencil for solder paste deposition to PCB's soldering pads are mostly manufactured from metal materials, which are shown at Table 1.

Table 1 Stencil estimated lifetime by material [2], [4].

Stencil material	Stencil lifetime (number of depositions)
Indurate nickel	100 000
Stainless steel	50 000
Standard nickel	50 000
Copper alloy	10 000
Soft nickel	10 000

The result from Table 1 is that the longest lifetime has indurate nickel, which has the best resistance to abrasive and chemical processes. Its high price include expensive electroform manufacture technology. For small series, where high accuracy is not requiring, is possible use a polyamide of polyimide stencil. But non-metal stencils have sort lifetime and they can have lower accuracy of deposition. There are three common used manufacturing technologies for metal stencils:

- chemical etching,
- laser cut,
- electroforming etching.

Main difference between these manufacture technologies is how small solder paste volumes can by possible deposited on PCB. Stencil made by chemical etching can be used for pith greater than 0,6 mm [5]. These stencils are relative cheep, but they have sort lifetime (cooper alloy), shape of stencil apertures has lower accuracy (Fig. 2).
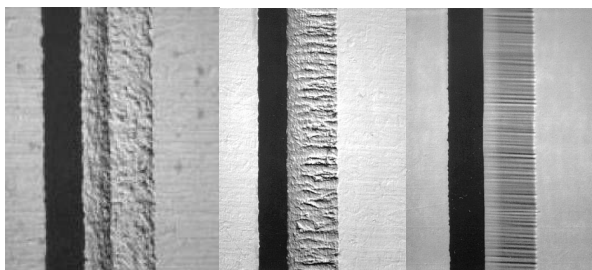
Fig. 2. Comparison of aperture walls surfaces made by chemical etching (left), Laser cut (middle) and electroforming etching (right) [4], [5].

Laser cut stencils are massive apply in industry because their good "value for money". Laser cut stencils can by use for components with pitch greater than 0,4 mm [5]. Manufacturing of these type of stencil is relative quick and also redesign is possible (add new or lager apertures). Advance is trapezoidal shape of apertures for smoother release of solder paste. The most expensive manufacturing technology for stencils is electroforming etching. Electroformed stencils have the smoothes wall surfaces of apertures (Fig. 2), for good solder paste release, the longest lifetime, because indurate nickel has good durability and also nickel has lower surface friction. This stencil can by use for ultra fine pitch components (> 0,2 mm [5]) like µBGA and Chip Scale Packages.

III. TRANSFER EFFICIENCY AND STENCIL DESIGN

Assemblies with miniature components such as 0201's or 01005's, CSPs or µBGAs packages and standard larger SMT components are challenging products. Problem is that, design a stencil that will reliably print the small solder paste volumes for the miniature components while providing sufficient solder paste volume for the larger SMT components and through hole components.

Stencil design is probably the most important of the process issues. Poor stencil design can lead to insufficient solder deposits, which can prevent the component from touching the paste or not have enough flux to overcome the oxide on the sphere or in the paste. Area ratio, as well as transfer efficiency, plays a huge role here. Although stencils may offer a small increase in the amount of paste applied, it is the paste itself that usually makes the difference. By feeding the stencil details into the paste measurement system at onset, the system can calculate the theoretical amount of paste that should be deposited, and then create a percentage (efficiency) by measuring the amount of paste that was actually deposited. Inconsistent or reduced transfer efficiency will reduce the total solder paste volume deposited on the solder pad and thus, possibly contribute to the reduced wetting of the flux, causing a head-in-pillow defect. Some variables that can affect transfer efficiency are stencil type, atmospheric conditions and the paste itself.

The biggest effect on transfer efficiency is stencil design and type of manufacturing. In generally, good solder paste deposition is when transfer efficiency is higher than 70%. For reach this number has to by articles fulfilled of Area Ratio limits. The Area Ratio is the ratio between the area of the aperture opening and the area of the aperture walls. When lower values of Area Ratio are find, it can by chose stencil manufacturing technology from Table 2.

Table 2. Choosing right stencil type by parameter Area Ratio.

Area Ratio	Recommended stencil type
over 0,90	chemical etching
0,90 – 0,66	laser cut
0,66 – 0,50	electroforming etching
under 0,50	necessary redesign of stencil

IV. RESULTS

This paper confirms that the electroformed stencil provided better print performance than the laser cut stencil or chemical etching stencil. Mirror aperture walls provided by electroforming stencil process, release solder paste better than the other stencils types. But because high price of electroformed stencil can be expensive use only electroformed stencils. For good "value for money" is necessary know the Area Ratio parameter in stencil design and determinate for correct stencil type.

REFERENCES

[1] PIETRIKOVÁ, A., ĎURIŠIN, J., MACH, P.: "*Diagnostika a optimalizácia použitia ekologických materiálov pre vodivé spájanie v elektronike*", ISBN 978-80-553-0447-2, Košice 2010

[2] SZENDIUCH, I.: "Základy technológie mikroelektronických obvodov a systémov", ISBN 80-214-3292-6, Brno 2006

[3] JIANBIAO P., TONKAY G., STORER R. H., SALLADE R. M., LEANDRI D. J.: "*Critical Variables of Solder Paste Stencil Printing for Micro-BGA and Fine-Pitch QFP*", IEEE Transactions on electronics packaging manufacturing, Vol. 27, No. 2, April 2004

[4] JOHNSON, A. - BOYES, B.: "*Stencil Printing Basics: Equipment Materials and Process Requirements*", Speedline Technologies Inc., 2002

[5] MALLIK, S., EREKE, N.N., DURAIRAJ, R., MARKS, A.E., SEMAN, A.: "*Wall-slip effects in SnAgCu solder pastes used in electronics assembly applications*", Materials & Design, Vol. 30, No.10, December 2009

14

Technical and economical proposal for the production of car batteries.

Gabriela Spišáková

Supervisor of bachelor work: prof. Ing. Irena Kováčová, CSc.

Department of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

gabriela.spisakova@student.tuke.sk

Abstract - **The subject of this thesis is the characteristic the production system, product and proper management of the production. There is pay attention to the production profile and the program production and also opportunities for proper handling with production means. The work refers to the importance of respect for the individual steps as well as the use of different calculation systems.**

Keywords - **production, manufacturing process, product, car batteries, stocks**

I. INTRODUCTION

The main and only subject of this production is the production of car batteries. There is only one variant of this product. It is called lead battery. It is actually a type of standard car batteries. The entire production process of the product has three phases. The first phase is deals with the production of basic parts. The second includes basic installation and fitment. The results of the third phase are finished products. In this case the car batteries.

II. PRODUCTION AND ITS STRUCTURE

Production of car batteries is destined for the market and it is sequenced among the mechanical production. Although that is produced only one product is manufacture of car batteries mass production. It is actually the mass production of one type of product in large quantities, practiced for a long time. In this production are applied special machines and lines, and human work is minimal. This production is typically organized as a continuous production.

A. Profile of the product and production

Car battery or the accumulators are also sources of electricity in motor vehicles. They are actually the sources of chemical energy, which is produced only when is the engine in the rest. In another case, the energy source is an alternator or dynamo. The battery is composed of several articles. Each article must have the approximate voltage 2.1 V. Very substantial part of the battery is a plastic container, in which are located all other components. Another part is the electrolyte separator. In the car battery has an important role the

lead. Lead allows you to add a great power. This fact is important when starting. Autobatteries suits well their frequent use. It is necessary to be cautious in their assembly and disassembly. At removal must first be disconnected skeleton outlet. At installation it is the opposite. Car battery must be recharged in a ventilated area, sources of direct current, and the same voltage as the battery. Battery terminals should be clean and well-tightened clamps. The average life of car batteries is 4-6 years. However, depleted below 1.75 V is irreversibly damaged. [1]

B. Program of production

Car battery production is difficult process composed of several steps. The manufacturing process begins with the production of plastic containers and covers, mostly made of polypropylene. The next step is the production of lead plates or grid. Then followed by the insertion of a mixture of lead oxide. It is actually powdered lead and other materials such as sulphuric acid and water, which are applied to the grids. Other material made of powdered sulphate is added to the paste and applied to the negative plates. Positive and negative plates must be separated to prevent short-circuiting. Separators are thin sheets of porous, insulating material. Another part of the process is the pairing of positive and negative plates and separator. Then the battery is filled with electrolyte, which is a mixture of sulphuric acid and water. The final steps are connecting the cover, insulation inspection, charging. Battery terminals are connected to electricity and the battery is charging for several hours. After the charge is shifted to the line where appropriate, clean and are placed it on the labels [2].

Per year is produced approximately 1008 pieces which means that if it has 252 working days so is made of 4 batteries per day. For one year are consumed 8225.28 kilograms of lead and 2016 kg of sulphuric acid.

III. PREPARATION

A. Technological preparation of production

Technological preparation of production of this product is the very theme of the comprehensive, which includes a series of important steps. The first step is the workflow, which is the sequence of work activities, which are already described in the previous chapter.

Other steps include working instructions and cutting scheme, which is creates the appropriate software as OPTIMIC or ARCON. [3]

B. Design preparation of production

This kind of preparation production is related to product design and builds on it. Contains information common to that proposal. It resulted in drawings and the BOM components to determine which is also used software. The BOM are shown in the picture below.

Bill of material (BOM)							
Company:				Product: Car Battery 12V 70Ah	List:		-1-
					Letters:		-1-
No.	No. of parts	Title	No. of pieces	Material	Pure dimensions in mm	Quantity content	
					Thickness x Width x Length	1 pc	Product
1.	0001	Plastic container	1	Polypropylene			
2.	0011	Cover	1	Polypropylene			
3.	0102	Lead plate	25	Lead			
4.	05006	A mixture of lead oxide		Powdered lead, H2SO4, water			
5.	08775	Electrolyte		H2SO4, water			
6.	87674	Separator	1	Porous synthetic			
Prepared by		Checked:				Approved by:	

Fig.1 BOM components

C. Economic Preparation of Production

These are specific activities related to conversion costs and prices. Economics training includes:

- Conversion of the consumption of materials. It is implemented through the distribution of individual groups of materials and on the basis standardized rules and product characteristics. This conversion is realized in units. To produce a single car battery is coated with lead weighing 8.16 kg, 2 kg of H2SO4, and 0.5 kg polypropylene.
- Conversion of material consumption in spending. In this step is very important to compare prices of suppliers. Based on the above data, calculate material consumption in spending.
Purchase price:
Lead - € 1.300 / tone, polypropylene-€ 1.052 / tone, H2SO4 - 1.43 € / kg.
Price of material to produce one piece of product is:
Lead - € 10.608
Polypropylene - € 0.52
H2SO4 - € 2.86
final price is: 13.988 €.
- Conversion of wage expenditure
- Conversion rates. It is indicated on the basis the calculation formula, which contains the sum of

factors: direct costs + direct labour + other direct costs + administrative expenses + selling expenses + profit. For this production is adapted to calculate one piece.

$$15$$
$$+34,23$$
$$+3,33$$
$$+4$$
$$+2,82$$
$$\underline{+6,3}$$
$$62,38 eur$$

DISPOSITION OF MATERIAL. It is a short-term emergency plan consisting of plans, the consumption, stocks and supplies and the ongoing continuous registration of consumption, inventory records and performance records of supplies. The essence of determining the consumption of material is its standardization. [4]

IV. PRODUCTION CAPACITY

It is the ability of the unit or system to carry out their activities in a certain period. In other words, this is the maximum output by the manufacturing unit will produce over time. Capacity is dependent on the technical level of production machinery and equipment, at the time of their activities, work organization and production, on manpower and materials. To calculate capacity are three types of formulas. The proposed company produces only one type of product capacity is expressed in units of kind and therefore the most appropriate relations for the calculation is the following relation:

$$Q_p = T_p x V_p \qquad (2)$$

where:
Qp - is capacity expressed in specific units kind,
Tp - is usable time fund in hours,
Vp - is kind exercise in units per hour.

Example:

Operation produces car batteries and used 14 machines (Pressure Dia Casting Machine, Battery Oxide Filling Machine, Oxide Mill System, Paste Mixing Machine, Automatic Grid Pasting Machine, Complex Alloy Casting Machine, Automatic Short Circuit Testing Machine, Fully automatic VRLA battery water cleaning and drying machine, Car battery automatic assembly line, Automatic Code Printing Machine, Automatic Interior Resistance Testing Machine, Automatic Leak Testing Machine, Automatic Aluminium Foil Sealing Machine, Automatic Heat Sealing Machine) and their parameters are given in the technical descriptions of equipment. The available fund one time the machine is 5193.18 hours per year. Usable time the fund is calculated by the equation: Fund calendar time - public holidays, weekends, holidays, leave, what time we reach

the nominal fund. After deduction of days for maintenance, repairs get usable result time fund [6].

Based on the data it is possible to calculate individual capacity.

Pressure Dia Casting Machine:

$T_p = 5193,18h$

$V_p = 120ks/h$

$Q_p = 623181,6ks$

Battery Oxide Filling Machine:

$T_p = 5193,18h$

$V_p = 180ks/h$

$Q_p = 934740ks$

Paste Mixing Machine:

$T_p = 5193,18h$

$V_p = 120ks/h$

$Q_p = 623181,6ks$

Complex Alloy Casting Machine:

$T_p = 5193,18h$

$V_p = 480ks/h$

$Q_p = 2492726,4ks$

Car battery automatic assembly line:

$T_p = 5193,18h$

$V_p = 100ks/h$

$Q_p = 519318ks$

Automatic Interior Resistance Testing Machine

$T_p = 5193,18h$

$V_p = 100ks/h$

$Q_p = 519318ks$

At following equipment is identical with the final capacity.

Is also very important to establish the capacity of the production areas.

$$Q_p = \frac{M}{m} \, x \, \frac{T_p}{d_v} \tag{3}$$

where:
M - is total production area in m²,
m - is capacity standard surface to produce a product in m²,
d$_v$ - is time interim standardized product in the production of 1 hour.

Example:
M = 750 m², designed on the basis of land-use decision
m = 4m²,
d = 0.0081 hour, calculated as the average of all the machines.

$$Q_p = \frac{750}{4} \times \frac{5193,18}{0,0081} = 120212500$$

Scope of capacity utilization is influenced by many factors and mainly the production plan, which determines planned capacity utilization.

The best way, which can increase capacity utilization to improve the organization of work, more use of working hours, etc. There are many ways to increase capacity and among the major ones include shortening operating times, trained workforce, and reduce labour content products [4].

V. PRODUCTIVITY

It concerns to all businesses. The ratio between the volumes of production inputs for the same period is determined by the ratio of productivity.

A. Labour productivity

Partial productivity p is determined by the same input ratio:

$$p = \frac{output}{input} \tag{4}$$

According to data in the table the productivity is determined for the previous year:

Inputs - outputs:			Year XXX
Outputs	products	Pieces Price EUR /pc	1008 62.38
Inputs	material	kg Costs EUR / kg	10745,28 1.26
	energy	kWh eur/kWh	300000 0.035
	work	Hrs. EUR / hour	6048 2.02

Tab.1.Table to finding productivity

$$p = \frac{62879,04 - (13539,0528 + 10500)}{12216,96} = 3,17$$

B. Total productivity

It is a summary of the productivity factor and expresses the resulting total efficiency of all resources.

$$C_p = \frac{v\acute{y}stup}{amount of \; resource \; inputs} =$$

$$= \frac{v\acute{y}stup}{work + capital + energy + material}. \tag{5}$$

$$C_p = \frac{62879,04}{13539,052 + 10500 + 12216,96} = 1,7343$$

VI. STORES

As with most in this production is necessary to determine the exact amount of inventory in stock and can also derive their functions. In this regard, here are three types of inventory, whose determination is very important. The first production inventory is which is actually supplying all the materials from suppliers. Inventories are stocks of production created their own materials that are produced in the previous phase and also preparations, which are delivered within the cooperative relations in one company. Equally important is the last inventory of finished products, which is the inventory-finished production. The purpose of this inventory is to ensure a smooth and trouble-free issue inventory items for consumption. According to the

function of individual inventory is very important to their classification. This classification is in the manufacture of tactical management. This classification is mainly ordinary inventory of the inventory, which is covering the needs in the period between two supplies. Further is an insurance inventory, which is very important as it is covering the bulk of deviation from planned consumption. It is necessary to define precisely the technical and seasonal inventory. As in every production and there is ultimately necessary to identify as important to smaller stocks in particular emergency, maximum, minimum, order and unused inventory. [4]

It is necessary that at each production standards are maintained inventory, based on parameters that can be precisely defined. One of these parameters is the supply cycle which is the time elapsed between two successive deliveries. This time is expressed in days. Another parameter is size of the delivery. It's height at once delivered quantity of material. It is expressed in physical units. Consumption is determined on the basis of actual consumption for some time and planned consumption. It is expressed in the quantity or monetary units. The last parameter is delivery time and ordering time. In this area the focus of the standardization of insurance inventory, which can ensure the correct determination of the high degree of reliability of material consumption in the enterprise, even in very difficult conditions. The basic way it is possible to calculate the insurance reserve is based on the number of days required to produce supplies + transfer to the supplier + implementation of orders + Transport from supplier to customer + take delivery in the enterprise + casual preparation before putting into consumption. Obtained insurance stock in days converted at the average daily consumption of buffer stocks in material units. There are many methods to calculate insurance reserves. In this case, is probably the most appropriate method using method of coefficient Protection class to which a fundamental relation:

$$Z_p = M_{pl} \cdot k_j \qquad (6)$$

where:
Mpl - is projected yearly consumption,
kj - is coefficient of insurance.

The idea is to find the coefficient of Protection class of the table. It is still created for a specific method. Searching for the basis points set for the material on the basis of the criteria that determine the supply of insurance and the requirements of their insurance. Buffer stocks are obvious to any data item of material. Fuse is important for items affecting consumption patterns, the lack of which causes defects in production.
M$_{pl}$ = 10745.28 kg
k$_j$ = 0.028

$$Z_p = 10745,28 \cdot 0,028 = 300,86 kg$$

As in most, and in this production inventory management succession management activities whose purpose is to find and secure the amount of stock of each species so that the material secured was continuous production process. It is obvious that this level of control is influenced by external but also internal factors. It is very important to be accurately set and monitor the status of orders, their terms, height and

extent of their storage. Management systems stemming from inventory, costs and organization can describe its own inventory management: a single ordering system, re-ordering system. It is necessary to determine the amount of suitable low cost associated with ordering and storage. Smaller amounts are equal to more orders and higher costs and the number increases the costs of storage. The average stock on the stock is defined by:

$$Z = Z_p + \frac{X}{2}$$

Then for the case:

$$Z_p = 300,86 kg$$

$$X = 11000 kg$$

$$Z = 300,86 + \frac{11000}{2} = 5800,86$$

It is also important that the designated storage costs without insurance stocks:

$$N_S = \frac{X}{2} \cdot S \cdot \frac{P}{100},$$
$$(8)$$

$$X = 11000 kg \sqrt{2}$$

$$S = 14099,904 eur$$

$$P = 20$$

$$N_S = 15509894 eur$$

And finally, the cost of ordering:

$$N_O = E \cdot \frac{M}{X}$$

$$(9)$$

$$X = 11000 kg$$

$$M = 10745,28 kg$$

$$E = 4000 eur$$

$$N_0 = 3907,37 eur$$

All calculations are indicated for annual consumption. The aim of this whole process is to minimize costs [5].

CONCLUSION

Production of car batteries is a complex process comprising a number of important tasks, which may reach incorrect compliance with the negative result. Like any production, this one has many negative effects, on which is paying particular environment. Despite this fact, today's mobilized world do without this product, which is part of almost any vehicle.

ACKNOWLEDGEMENT

REFERENCES

[1] http://cs.wikipedia.org/wiki/Autobaterie

[2] http://www.batterycouncil.org/LeadAcidBatteries/ HowaBatteryisMade/tabid/107/Default.aspx

[3] http://www.skolasnov.ic.cz/matury/technicka_priprava_vyroby __master.pdf

[4] http://www.comeqmachinery.com/Products/ BatteryRelatedMachinery/AutomaticHeatSealingMachine/index. html

[5] Synek M., Manažérska ekonomika, Praha: Grada Publishing, 2007, ISBN 978-80-247-1992-4

[6] http://www.euroekonom.sk/download2/maturitne-otazky/Ekonomika-Maturitna-otazka-EKO_06.pdf

[7] http://www.podnikam.webnoviny.sk/marketing-a-reklama/ake-naklady-vstupuju-do-ceny-vyrobku-kalkulacny-vzorec-ktory-ulahci-stanovit-cenu-vyrobku-/14973

14

The measuring stand for testing of basic circuits with operational amplifiers distributing through internet

[1]Marek GODLA, [2]Martin SEKERÁK

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

marek.godla, martin.sekerak@tuke.sk

Abstract-- This project was created for effective studying practical sections of object „Basics of electronics" by a real remote control measuring workplace. The content of this measuring workplace are circuits of operational amplifiers and measure instruments and web camera. In circuits, it is possible to change a value of feedback resistors and setting parameters of measure instruments by PC. The measuring workplace is distributed trough the internet. The whole project is created in the development environment LABview with using the measuring card NI- DAQmx.

Keywords-- Labview, remote control measure workplace, operation amplifier, feedback resistors, remote control measure instrument, NI-DAQmx.

I. INTRODUCTION

Testing of basic circuits with operational amplifier for students is important to understanding of elementary electronic. Study of electronic without real measuring and exploring features of basic electronics components is not sufficient for students. A solution of this problem is real measuring in a laboratory. However this measure is problematic because of inadequate number of measuring instruments and fast destruction of measuring components. The destruction of electronics components is caused by often using.

A main task of this work is to design a measuring workplace, which can be controlled by a user via internet. The program which controls of circuit for different laboratory tasks through Internet was developed in the LabVIEW. LabVIEW development environment allows to remote access from any client with available web browser like Explore, Mozilla. In this workplace is used real instruments controlled from the program using their interfaces.

II. DESIGN AND CREATE MESASURING PLANT

Laboratory stand consist of four independent circuits for four laboratory exercises. Each measuring circuits is being controlled by a switching circuit which allow to change circuit component according command by a student. The first two measuring circuits contain operational amplifiers connected as inverting and noninverting amplifier, where the feedback resistances were controlled remotely by user. Change of feedback elements caused different characteristic of amplifier. Final value of feedback resistor is determined by parallel connection of four resistors connected through the analog multiplexer (fig. 1). Third measuring circuit represents application of an operational amplifier as integrator. Fourth circuit to be prepared for laboratory measurement was the rectifier with operational amplifier. The input signal from generator is being connected to all four circuits parallel.
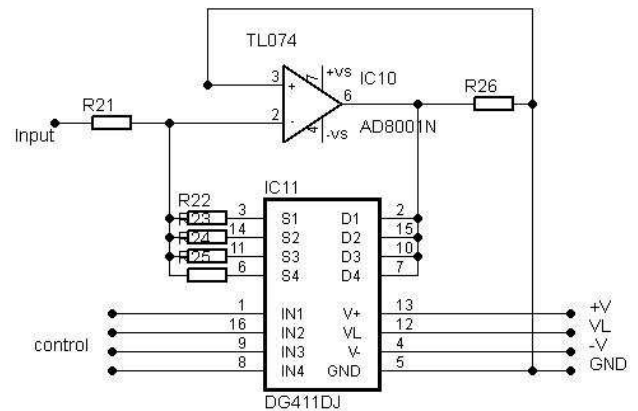
Fig.1 Switching of feedback resistor on inverting circuits of operational amplifier

Every there are two measuring nodes in any circuit . Each measuring node is connected to one of eight inputs at analog multiplexer. The Output from the multiplexer are connected to input of stand alone oscilloscope. Second input of oscilloscope is connected to the output of generator. Selection of measured nodes by any multiplexer is performed independently by 3-bit addressing bus AB. The feedback resistors in both amplifiers and rectifier and feedback capacitors in the integrator are chosen simultaneously by control bus CB (fig.2). Multiplexers for control of the feedback resistors and selection of signals from measuring nodes are realized by integrated circuits DG 408, 409 and DG 412. The ON resistance DG 408, 409 is 100Ω and DG 411 is 35Ω. Multiplexers are controlled by TTL/CMOS outputs from

measuring card NI-PCI 6251. It contains as well as a general breadboard area and power bus terminals for external power and grounding. Schematic of whole plant is on (fig 2). Driving software use basic NI-DAQM blocks: open session, write, read, and clear session. Software is orientated to easily controlling by user.

USB port was utilized for communication between server and instruments. Control commands were generated by LabVIEW using VISA language and standardized SCPI commands.
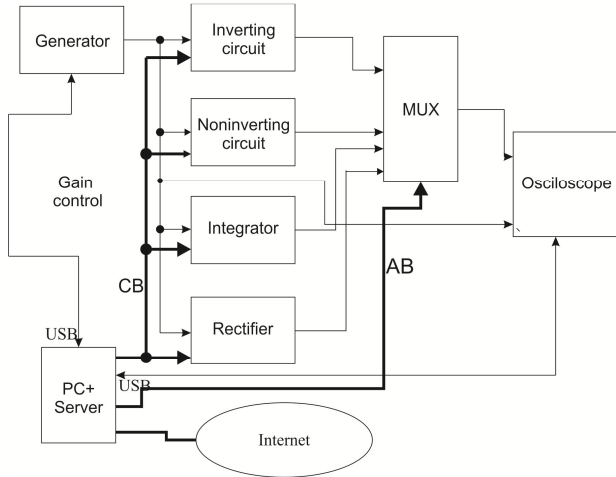
Fig.2 Block scheme of laboratory stand

IV. DESIGNING A WEB SERVER FOR CONNECTION A WORKPLACE TO THE INTERNET.

The web server is designed in the environment LabVIEW. The LabVIEW environment allows to generate the html web page from the panel diagram of the control program which is accessible from the client´s computer. Program is running on server which is able to control laboratory stand. Clients computer achieves access to the server enabling Remote Panel Server access in the control program at the server. User from the clients computer can control program using Explorer or Mozilla. These web browsers have to allow pop-up application.

The link of this web page is set on the official web pages of the laboratory Department of electronics and multimedial communication.

V. CONTROL PROGRAM FOR CONTROL MEASURING INSTRUMENTS

Control program allows changing basic setting-up of generator AGILENT 33220A and oscilloscope RIGOL DS1052E. Generator AGILENT 33220A represents arbitrary signal generator of frequency range 0-20 MHz. Oscilloscope RIGOL DS1052E is two channel digital oscilloscope with max 1GSa/s.

Multiplexing circuits for connection of selected measuring points into two inputs of oscilloscope were controlled from the digital outputs of the USB DAQ board NI6251 by National Instruments. For any multiplexer three digital outputs were utilized. Other four digital outputs were used for control of combination feedback resistors simultaneously in

all four circuits. Four digital inputs allow to set-up one of 16 resulting values of feedback resistor. Commands for digital control of laboratory stand – its all four circuits- were generated from LabVIEW control panel.
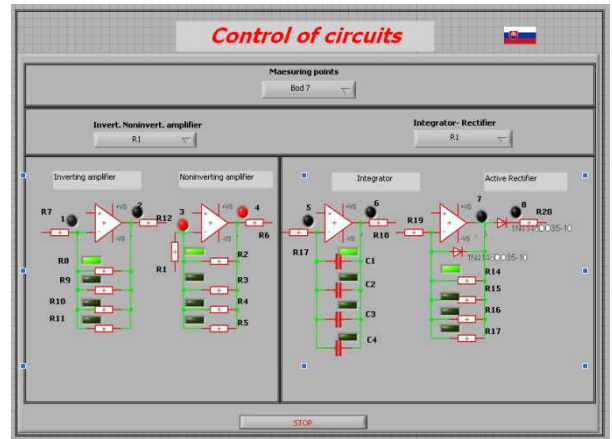
Fig.3 Control program for switching of feedback resistor

Both instruments were controlled via USB link and using VISA standard.

Control program of oscilloscope allows controlling basic settings:
- time base
- voltage range
- ON/OFF inputs
- triggering level
- V and H -position
- DC or AC coupling of channels
- Switching on/off AUTO button and RUN/STOP buttom

Moreover program allows acquiring following parameters of the measured signals:
- measured signals from two input channels
- transferring values from oscilloscope determining parameters of measuring waves (amplitude, period, minimal voltage, maximal voltage, frequency)

Control program of generator allows controlling basic settings:
- signal form
- signal amplitude, phase, offset, and frequency
- output polarity
- ON/OFF channel

Panels which control generator, oscilloscope and measuring nodes are on fig.4,5
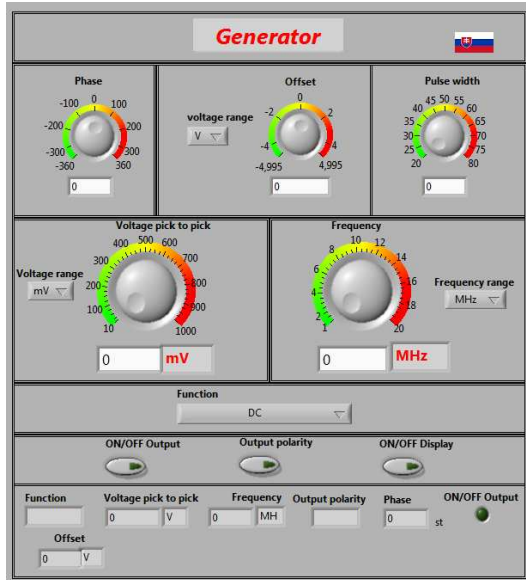
Fig.4   Control panel of generator
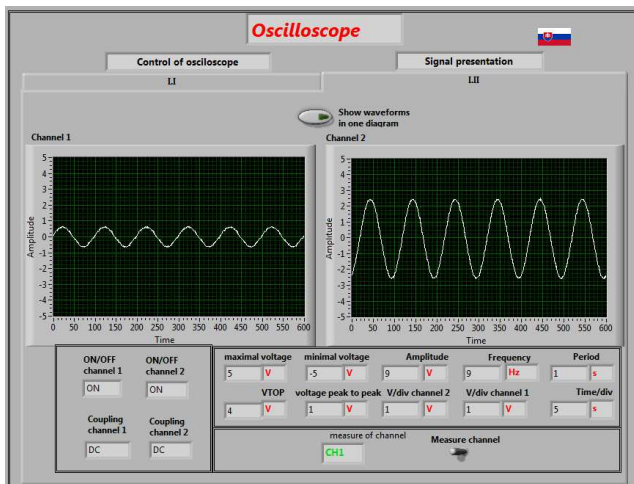
Fig.5   Control panel of oscilloscope

VII. CONCLUSION

Presented laboratory stand controlled remotely through Internet serves students of Faculty of Electrotechnic and Informatics for distant laboratory exercises in the course of "Basic electronics". Described stand is one of four stands which are in the trial phase. Three laboratory stands are controlled with DAQ board only. Described one has to increase feeling of work in the real laboratory. Results from questionnaire inquire among students showed that both methods how to control of laboratory stands are for users similar. In the future this stand can be enriched with a web camera and some more circuits with an operational amplifier.

This article was created from my Diploma work [7]. Whole project was running in laboratory on department KEMT of FEI TUKE. Project gives a new idea about measuring via the internet and proposal of effective study of electronics.

Issue of this project doesn't have any conjunction with my PhD study and my future PhD thesis.

REFERENCES

[1] www.ni.com
[2] RIGOL, "Programming guide DG 1022 Function/Arbitrary Waveform Generator, RIGOL Technologies Inc., August 2009
[3] RIGOL ,"Programming guide DS1000, DS1000CA Series Digital Oscilloscope, Publication number: PGA02100-1210, RIGOL Technologies Inc, June 2008
[4] USER GUIDE SCC- 68 ,"I/O Connector Block with 4 SCC Signal Conditioning Slots for DAQ Devices, National Instrumnets Corporation, 2006
[5] Data Sheet, "DG408, DG409", File Number 3283.5, Intersil Corporation, June 1999
[6] Data Sheet, "TL074, TL074A – TL074B", STMicroelectronics Group of Companies, December 1998
[7] Godla, M.: "Diploma Work", Technical University in Košice, Faculty of Electrical Engineering and Informatics, Slovak Republic, 2010

[8] FINKA Filip, "Diploma Work", Technical University in Košice, Faculty of Electrical Engineering and Informatics, Slovak Republic, 2008
[9] HORVATH Juraj, "Bachelor Work", Technical University in Košice, Faculty of Electrical Engineering and Informatics, Slovak Republic, 2008

14

The Ultrasonic Wirebonding in Electronics

[1]Igor VEHEC, [2]Michal KRAVČÍK, [3]JurajĎURIŠIN

Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic
[2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]i.vehec@tuke.sk, [2]michal.kravcik@tuke.sk, [2]juraj.durisin@tuke.sk

Abstract—**The paper briefly deal with wire-bonding and factors that affects on the wire-bonded joints. The wire-bonded joints are widely used in a lot of applications where is they quality very important. The quality of wire-bond joint is affected by the many factors as materials, method of wire-bonding, parameters of wire-bonding etc.**

Keywords—**wire-bonding, quality of wire-bonded joints.**

I. INTRODUCTION

The wire-bonding is one of the technologies for interconnecting of the chip with external environment and it is one of the critical steps in the completion of the components. The requirement for quality of wire-bonding process (and joints) is all the more important, that the many demanding applications consist from microchips (among others components) with electrical interconnection based on wire-bonded joints. Creating of quality wire-bonded joints is a precondition for adequate reliability of electronic components, particularly in critical and demanding conditions.

II. WIRE-BONDED JOINTS

The wire-bonded joints can be divided either according to physical principle on which they were bonded, or according to the shape of the wire bond. From a physical principle distinguish thermocompress wire-bonding (T/C), ultrasonic wire-bonding (U/S) or thermosonic wire-bonding (T/S) (Tab. 1). In term of shape of bond, we distinguish the ball bonding and wedge bonding (Fig. 1 and Tab. 2).

The T/C bonding technology is the oldest method for creating of wire-bonded joints. The main disadvantage of T/C bonding is the relatively high temperature and high force during the bonding, which leads to significant problems during thinner chips bonding. The other two methods of wire-bonding (U/S and T/S) are more suitable technologies for wire-bonding, which offer several advantages [8].

The modern U/S wire-bonder equipments use a typical frequency in the range of 120 kHz to 140 kHz, while the older equipments work with a frequency of 60 kHz. The transducer electrical energy to mechanical energy is usually based on piezoelectric element. Amplitude of the generated ultrasonic waves is subsequently amplified by the element called "horn" (this element narrows their diameter) with the "tip" (bonding tool) at the end [1], [7].

The wedge bonding (T/C since 1957, U/S since 1960 and T/S since 1970) is older as ball bonding technology. The ball bonding is faster and more robust methods, bud in spite of this, the wedge bonding dominates in cases where they are requirements on low profiled interconnection, uninterrupted stitch or bonding chip/chip.

The basic adjustable parameters of wire-bonding are (1) force of wedge applied on the contact pad during bonding, ultrasonic (2) power and application (3) time of the ultrasonic energy. The precise setting of parameters depends e.g. on used materials, shape of "tip" (not only ball/wedge), and greatly affects the quality of wire-bonded joint. The shape and

TABLE I
DIFFERENT TYPES OF WIRE-bonding [3], [9]

	Wire-bonding process		
	T/C	U/S	T/S
Force	high	low	low
Temperature	300 – 500ºC	25ºC	100 – 150ºC
Ultrasonic energy	no	yes	yes
Wire material	Au	Au, Al	Au
Pad material	Al, Au	Al, Au	Al, Au

TABLE II
BALL VS. WEDGE WIRE-BONDING [3], [8], [9]

	Ball wire-bonding	Wedge wire-bonding
Wire-bonding	T/C, T/S	T/S, U/S
Wire material	Au	Au, Al
Bonding tool	capillary	wedge
Bonding tool material	tungsten carbide ceramic	Tungsten carbide – for Al Titan carbide – for Au
Speed	10/sec (T/S)	4/sec

quality of wire-bonded joint are also affected by shape and parameters of used bonding tool (tip), and on the other hand, type and shape of bonding tool depends on the specific requirements given by application. For example, the angle of wire under which enters to the bonding tool at wedge bonding may be different. Smaller angle (30°) is suitable for interconnections with request on higher loop, but wire is exposed to greater stress than at angle (60°), which is more suitable for the lower loops. For all that, the higher loops are more suitable in term of keeping of loop shape. [9]
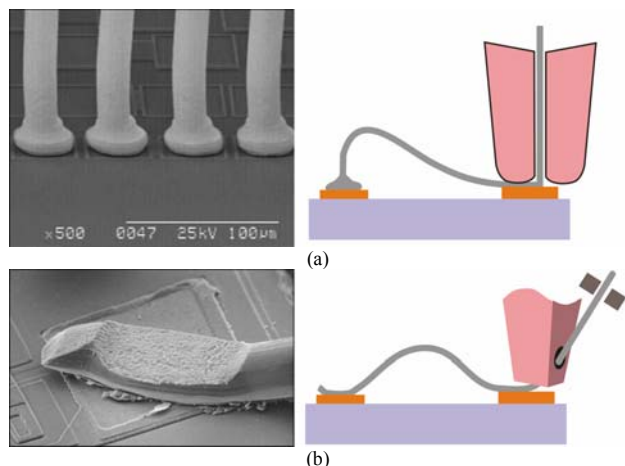
Fig. 1. Shape of ball (a) and wedge wire-bonded joint (b) [8].

III. MATERIALS USED IN WIRE-BONDING PROCESS

The bonding process is creation of conductive interconnection with the wire between bonding pad on the chip and bonding pad either another chip or substrate on which is chip placed. Thus, there are two major materials: (1) material of wire and (2) material of pad, in the whole bonding process. The properties of used materials as well as compatibility between them are other significant factors that affect quality of wire-bonded joint.

A. Materials for wires

The wires material based on Au, Al or Cu are the most common used materials for wire-bonding in electronics. Each of these materials has specific properties that determine the advantages and disadvantages in terms of their usability in wire-bonding process.

Au wires are widely used for T/C and T/S bonding. Ultrapure Au is too soft and therefore there are added ingredients such as Be (5 to 10 ppm) or Cu (30 to 100 ppm). There is required higher temperature (typically 150°C) when are Au wires wire-bonded [3], [8].

Au wires are in its pure form very soft too, so Au wires are most often used in the form of alloys (1% Si or Mg). There is do not required higher temperature during wire-bonding process [3], [8].

Cu based wires appear to be a suitable replacement of Au wires for economic reasons. Cu offers significantly better conductivity and higher fusing (melting) current, but it is easily oxidize and it is worse bondable [3], [6].

Wire thickness depends on application requirements and varies from ~10 μm to 100 μm. The thinner wires are suitable for fine applications. The thicker wires have higher strength and better resistance to deformation, which can occur during packaging [9]. They are also suitable for applications which require higher current loads.

B. Compatibility between wire and pad materials

The typical materials for pads are Au, Al, Cu, Ag or Ni. The wire-bonded joints based on equal materials for wire and pads are least susceptible to intermetallic compound forming and therefore they are highly reliable.

The combination wire/pads based on Au/Al or Al/Ag are vulnerable to formation of many intermetallic compounds and

Kirkendall's voids. Intermetallic compound $AuAl_2$ is formed at the interface Au/Al even during wire-bonding at room temperatures. Nevertheless, Au/Al is one of the most common used combinations. Al/Ag combination is widely used in the field of hybrid integrated circuits, because used thick-film pastes are common based on AgPt or AgPd. Joints tend to degrade and oxidize in a humid environment.

The combination wire/pads based on Au/Cu or Cu/Al are vulnerable to formation of many intermetallic compounds. Wire-bonding of Cu/Al is characterized by the formation of that leads to similar defects as the system Au/Al, but growth of intermetallic compounds is slower and without Kirkendall's voids.

The combination wire/pads based on Au/Ag or Al/Ni are more reliable as previous mentioned combinations. Au/Ag are characterized by high reliability for long periods even at elevated temperatures, because they are not prone to the growth of intermetallic compounds. Al/Ni is typically used for power components and for high-temperature applications and it is more reliable as Al/Ag or Al/Au combination in various operating environments.

IV. ASSESSMENT OF WIRE-BONDS QUALITY

Optimization of wire-bonding process requires considering number of factors that affect the quality of wire-bonded joints (chip – package – wire – pads, etc.). There was proposed several methods (destructives or nondestructive) for assessment of wire-bond joints. Pull test and shear test are standard destructive methods for assessment of the mechanical properties of joints. Pull test is one of the widely accepted methods of testing. Visual inspection, nondestructive pull test (NDPT) and nondestructive shear test (NDBS) are typical nondestructive tests [4], [7]. Overview of standards used for assessing of wire-bond joints are summarized in Tab.3 [3][10].

The pull test is one of the appropriate tests for assessing of wire-bond joints, which partly correlates with visual

TABLE III
PREVIEW OF STANDARDS USED FOR ASSESSING OF WIRE-bond JOINTS [3], [8], [9]

Standard	Name/Description
EIA*/JESD22-B116	Wire Bond Shear Test
ASTM** F 458	Standard Practice for Nondestructive Pull Testing of Wire Bonds
ASTM F 459	Standard Test Methods for Measuring Pull Strength of Microelectronic Wire Bonds
IPC-TM-650	2.4.42.3 - Wire Bond Pull Strength
MIL-STD-883H	Method 2010.12 - Internal Visual (Monolithic) Method 2011.8 - Bond Strength (Destructive Bond Pull Test) Method 2017.9 - Internal Visual (Hybrid) Method 2023.6 - Nondestructive Bond Pull Method 3003.1 - Delay Measurements Method 2001.3 - Constant Acceleration Method 2026 - Random Vibration Method 2002.5 - Mechanical Shock Method 1008.2 - Stabilization Bake Method 1004.7 - Moisture Resistance Method 5009.1 - Destructive Physical Analysis

* Electronics Industries Alliance
** American Society For Testing And Materials

inspection as well as with the basic adjustable parameters of wire-bonding equipment (power, time and force). The

(a)

(b)
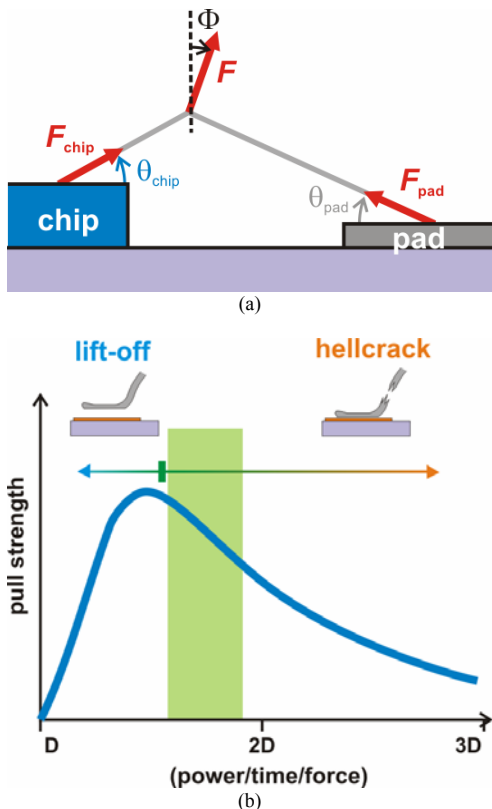
Fig. 2. Principle of pull test (a) and bond pull strength vs. deformation – or power/time/force (b) [8].

evaluation of the pull test is the one of the methods that allows optimize parameters of wire-bonding process.

The basic layout of the pull test is show in Fig.2a. Tensile force acts on wire and extends towards the two contacts (on chip and pads) during the pull test. In the general case, the force applied on the chip is [1], [11]:

$$F_{chip} = F\frac{\cos\left(\theta_{pad} - \Phi\right)}{\sin\left(\theta_{chip} + \theta_{pad}\right)}$$

and on the pads [1][11]:

$$F_{pad} = F\frac{\cos\left(\theta_{chip} + \Phi\right)}{\sin\left(\theta_{chip} + \theta_{pad}\right)}$$

Optimization of the wire-bonding parameters is determined based on the graph – maximum pull strength vs. power, time or force. All types of dependence exhibit similar trend and they correlate with the dependence of the maximum pull strength vs. wire-bond deformation (Fig.2b – the deformation of wire-bond is expressed as a wire-bond diameter D multiple). The wire-bond is stronger and more deformed while increasing any parameters of wire-bonding. The pull lift-off process of bond failure dominated at the beginning of the curve. The heelcrack process of bond failure is typical after reaching a maximum. The optimal parameters are located in a narrow range over the peak of the curve [3].

V. CONCLUSION

Assessment of quality of wire-bond joints is a complex process that requires extensive knowledge of the principle of the wire-bonding, mechanical and electrical properties of used materials with connections to the relevant norms and standards. For the quality wire-bond, it is necessary to take into account not only circuit layout and next technological processes but also supposed working conditions of components.

REFERENCES

[1] G. Harman, "Wire Bonding in Microelectronics", 3rd Edition, New York: McGraw-Hill, 2010, ISBN: 978-0-07-164265-1.
[2] *Bonding Wedge Catalog* (Book style). Kulicke & Soffa, 2001.
[3] Z. Lai and J. Liu, *Chapter A: Wire Bonding.* In: The Nordic Electronics Packaging Guideline, 2000. http://extra.ivf.se/ngl/
[4] C. Wang and R. Sun, *The Quality Test of Wire Bonding.* In: Modern Applied Science, Vol. 3, No. 12, Canadian Center of Science and Education 2009, ISSN 1913-1852.
[5] D. J. Beck and A. C. Perez, *Wire Bond Technology - The Great Debate: Ball vs. Wedge*, In: Advanced Packaging, March, 2007.
[6] Ch. Breach, *The Great Debate: Copper vs. Gold Ball Bonding*, In: Advanced Packaging, October, 2008.
[7] H.K. Charles, *Advanced Wire Bonding Technology: Materials, Methods, and Testing*, In: Materials for Advanced Packaging, Springer Science+Business Media, LLC 2009, ISBN 978-0-387-78219-5.
[8] Gaiser precision bonding tools, CoorsTek, Inc. http://www.gaisertool.com/
[9] Bonding Wedge Catalog, Kulicke & Soffa 2001. http://www.kns.com/
[10] MIL-STD-883H - Test Method Standard Microcircuits.
[11] Wire Bonding Quality Assurance and Testing Methods, http://ssd-rd.web.cern.ch/ssd-rd/bond/talks/2-04_Schneider-Ramelow.pdf

Utilization of Content Management System for Implementing the e-Learning Systems

Dominik Demeter, Juraj Banský

Department of Technologies in Electronics, FEI TU of Košice, Slovak Republic

dominik.demeter@tuke.sk, juraj.bansky@tuke.sk

Abstract— **In the article we deal with the main lines of e-learning system, that will be implemented at the Department of Technologies in Electronics and which is based on an open-source Content Management System (CMS). In the article are compared two most used CMS to find one, that will match our requirements or which should be easily modified to match them. Our goal is to define the main features of the e-learning system and find modules for CMS system, by which these features could be implemented.**

Keywords— **Blended-learning, CMS, Drupal, E-learning, LMS.**

I. INTRODUCTION

Nowadays, computers and information technologies are used in every area in our life to save a lot of time as well as money. This trend hit the educational systems, too.

Most of universities and their departments are using some kind of e-learning technologies or plan to implement them. In most cases, with expression e-learning, the universities mark classical study materials published in electronically form, what is not correct, but sometimes, we can see, that they have implemented a complex e-learning system.

This trend met our department, too. In next few years, we would like to implement a complex e-learning system, which will contain e-books, study materials in electronic form, on-line courses, interactive applications, animations, and the database of relevant publications, video-archive from lectures and from laboratory exercises, on-line exercises, tests and quizzes.

II. E-LEARNING

A. What is it

The e-learning expression can mark a simple electronic presentation as well as a complex learning system integrating on-line courses, study materials in electronic form, some kind of animations, applications or virtual laboratory allowing remote work on real technological equipments.

The noun e-learning can be defined as "learning using electronic: the acquisition of knowledge and skill using electronic technologies such as computer- and Internet-based courseware and local and wide area networks"[1].

A well-designed e-learning system should on the one side provide a complex, user-friendly environment for students to support their study, but on the other hand, it should not take more loads on teachers. It should provide an easy-to-use interface for creating study materials and on-line courses as well as a transparent navigation and administration interface.

B. Main facilities of an e-learning system

A well-designed e-learning system must have several critical facilities, without which we can not call a system as e-learning.

One of the most important features is the user management. E-learning system must be able to handle different user roles, e.g. anonymous user, student, teacher, course-administrator and main administrator. Every of these user roles must be configurable per user based so, that the system must allow to change permissions for every user, to restrict access and unauthorized operations.

The second important feature is content management. A quality of the e-learning system must support different kind of content. The most used types of contents are web-pages, e-books, on-line courses, quizzes, multimedia files (audio and video streams) and animations.

Connected with content management, it is very important, that an e-learning system must allow controlling the access of the each content. The access control can be user-based, based on date and time, on the satisfaction of some conditions (read the previous chapters, spent defined time on the previous page, accomplish some test, etc.).

A very useful feature should be the versioning of the content. If some changes occur in the study materials, the student can see the date and time of these changes, and is able to get back for the previous version.

If we are talking about content management, one of the critical tasks is to create a good index of content. E-learning system should have the ability to assign some tags for the each content, by which the content can be searched. Tags can belong in some categories, like courses, teachers, study areas, etc.

It is also very important to give the students a space for discussion. This discussion can be course-related, subject-related. The best solution is to create a discussion for every study material, where can the students exchange their experiences as well as they can ask the teacher some questions and where is some extra place for some auxiliary explanation of the topic of the study material.

A complex e-learning system should provide a space for exchanging files, what is important in the process of collecting and evaluating of student's works and what is a good solution for the problem of the file size limit of the e-mail attachments. There should be also a place for the communication related to exchange files, e.g. evaluation comments, hints, etc.

Crucial feature is the ability of assessment of the work that is done on the site – grade-book, test metrics, a cumulative total for test scores completed by a student during the course.

C. Blended learning

For universities, a typical e-learning system, which is the only education form, is not a desired solution. Emerging trend is to involve blended learning programs, in which, e-learning is integrated with traditional training methods. There are combined offline and online forms of learning, where the on-line learning usually means "over the Internet or intranet," and offline learning happens in a more traditional classroom setting. We assume that even the offline learning offerings are managed through an on-line learning system. An example of this type of blending may include a learning program that provides study materials and research resources over the Web while providing instructor-led, classroom training sessions as the main medium of instruction.

D. Learning management systems (LMS)

A LMS is a set of software tools, for delivering, tracking and managing on-line training and education. The main advantage of the LMS is the integrated environment.

Some LMS are web based to facilitate administration and access to learning content. Most of these systems are commercially developed and are provided under commercial software licenses, but there are several ones, that have open-source license. To implement an e-learning system at a university, better choice is to use an open-source alternative, which is providing in most cases every needed feature, and if not, they could be easily added on.

The most known and most used open-source LMS is Moodle. It is used at some departments at our university, too. As everything, an LMS system has their contras, too. Sometimes, it is hard to import some type of content; the user interface can not be modified; in Moodle there are implemented too many functions, and sometimes the whole system is too slow; it is not existing solution for atypical problems; the development community can help only with common troubles.

E. Sharable Content Object Reference Model (SCORM)

Connected to LMS, it is very important to mention SCORM, which "integrates a set of related technical standards, specifications and guidelines designed to meet Scorm's high-level requirements – accessible, inter-operable, durable and reusable content and systems". Content, that meets SCORM standards, can be delivered to students via any SCORM-compliant Learning Management system using the same version of SCORM with that is the content compatible.

It is very important to creating study materials that meets

SCORM standard, for using the materials any-time, anywhere, with any system. It is easier migrated the learning system to other LMS or to exchange learning materials between different learning organizations with no additional cost.

The support for SCORM is one of the crucial requirements for the CMS system that will be used for implementing e-learning system at our department.

TABLE I [2]
HIGH-LEVEL REQUIREMENTS FOR SCORM

Requirement	Explanation
Accessibility	The ability to locate and access instructional components from one remote location and deliver them to many other locations
Interoperability	The ability to take instructional components developed in one location with one set of tools or platforms and uses them in another location with a different set of tools or platforms.
Durability	The ability to withstand technology evolution and changes without costly redesign, reconfiguration or recoding.
Reusability	The ability to incorporate instructional components in multiple applications and contexts.

III. CONTENT MANAGEMENT SYSTEM (CMS)

A. What is it

CMS is a system used to manage the content of a web site. Usually, a CMS contains two main elements – the content management application and content delivery application. While the first element allows for a person, who do not know Hypertext Markup Language (HTML) to creating and managing content, the second element is responsible for updating and viewing the Web site.

The Top 3 most used CMS systems are Joomla, Drupal and Wordpress. Both, Drupal and Joomla are easily modifiable and have the support of a big developer community. The full source codes of every their modules are available and both have a very wide range of plug-in modules. To use one of them for our e-learning system should be a good bet. Wordpress is no as easily modifiable and have no plug-ins for extend the system for e-learning abilities.

B. Main integrated features of Drupal and Joomla required for e-learning system

Content management systems have some common features, like user management, multi-level permission content management with user-friendly interface, preview the content before publishing, templates, version management, forums, search engine, module for uploading files, polls and others. Every CMS have built-in support for plug-in modules, by which the features of the CMS system could be extended.

The built-in user management usually allows creating different user roles, by which the administrator can allow or restrict some operations for the user. Every CMS system has built-in support for anonymous user role. Usually, for every user or user-role, it is possible to define, which content is available and which modules they can use.

The most used content type, like classical web page, quiz, and flash animations are supported by default. Most of CMS

systems have built-in option for defining our own content type.

Drupal have in base configuration also the support of versioning. For the content it is possible to choose, if the older version of content should be saved, or overwritten. In case the versioning is selected, the date and time of the modification and the older content of the page is saved and can be showed for users with appropriate privileges.
Drupal have also integrated taxonomy – tagging system, which allows us to creating tags by vocabularies and terms, which could be used for searching the content. It allows classifying content into categories and subcategories.

A. Main integrated features of Drupal and Joomla required for e-learning system

The main power of the most used CMS systems are the wide range of plug-in modules, which are usually provided under open-source license and can be easily modified. There is also a big community of developers, who are ready to help, if the user/system administrator has some problem, which seems to be unsolvable. Usually, for the most of potential problems there is a solved user request with a short how to describing the way to solve it.

There are four main projects, which are based on Drupal and contain different tools helping educational process: Eduglu, ELMS, Open Scholar and Voice Box. These systems are containing a classical Drupal installation and contain also some extra plug-ins needed for educational purposes usually developed by some e-learning institute. It could be a good choice to start with some of these systems, because usually, those have integrated a lot of requested functions and the standard Drupal plug-in modules can be used with them, too. The combination of built-in modules and the wide range of additional plug-in modules give us a very powerful tool, which can solve all of our demands.

TABLE II
MODULES PROVIDING REQUESTED FEATURES

Requirement	Joomla	Drupal
Versioning	Simple Content Versioning	Built in
Discussion related to content	JComments	Comment and Tracker
Multimedia integration and player	All Videos	SWF Tools
User management	Artof User	Filter Permissions, User protect
File upload	Simple File Upload	Upload, File Field
Tags	Tag	Taxonomy
WYSIWYG editor	JCE	Wysiwyg
SCORM support	No free version available yet	CCK Scorm Field, Simple_LMS – under development
Gradebook	No free version available yet	Gradebook, School Administration
Searching	OpenSearch	Search, Custom Search,
Themes	Templates	Templates
Multilingual support	Translation Manager	Translation Management

There are also available complex learning systems based on Joomla, but they are usually commercial projects, so, it is not

possible to use them for free. This is the main reason, why I had chose the Drupal, as the base for our e-learning system.

IV. CONCLUSION

As it came out from the comparison of the two most used open-source CMS, Drupal seems to be the better choice for implementing an e-learning system, because it has more needed feature implemented and provided under free and open-source license. For every our requirement for the e-learning system, there are exists built-in or plug-in modules.

In the future, before implementing the e-learning system, it should be useful to compare Moodle and Drupal with appropriate plug-ins, to test, which one is more user-friendly, easier for administration and faster at the same environment. There are exists some projects, where developers are trying to connect Moodle and Drupal to a homogeneous system, which should inherit the best properties from both systems. If this project finish before we start implementing our e-learning system that could be a good choice, too.

REFERENCES

[1] MSN Encarta – dictionary, http://encarta.msn.com/
[2] SCORM Documents – What Is SCORM, http://www.adlnet.gov/Technologies/scorm/SCORMSDocuments/What%20Is%20SCORM.aspx.
[3] Harvi Singh, Chris Reed, A White Paper: Achieving Success with Blended Learning, Centra Software, 2001
[4] Drupal vs Joomla vs WordPress – Open Source CMS Comparison, http://www.comentum.com/drupal-vs-joomla-cms-comparison.html
[5] P.A.U. Education, S.L., eLearning Papers, 2007, ISSN 1887 – 1542.
[6] Ján Šúst, E-learning pro studenty se specifickými nároky, Masaryková univerzita Brno, 2006
[7] Official homepage of Moodle, http://moodle.org
[8] Official homepage of Drupal, http://drupal.org
[9] Official homepage of Joomla, http://www.joomla.org
[10] EFQUEL – European Foundation for Quality in E Learning, http://www.efquel.org/
[11] Dr. Amy Finn, Ph.D, Trends in e-learning, Centra Software, 2004
[12] Lepiš Fedor, E-learning – komunikácia zajtrajška, UMB SK
[13] Michal Novák, E-learning - nástroje pro tvorbu a řízení výuky, EF CZU, 2007

Vector control of induction motor with RF MRAS speed estimator

[1]Peter NGUYEN, [2]Matúš HRIC, [3]Karol KYSLAN

[123]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]peter.nguyen@tuke.sk, [2]matus.hric@tuke.sk, [3]karol.kyslan@tuke.sk

Abstract— **This paper deals with possibilities of sensorless vector control induction motor. For identification of the angular speed is used model reference adaptive system based on rotor flux (RF MRAS). This type of speed estimator is verified by simulation and some tests to changes of stator resistance are done.**

*Keywords—***induction motor, RF MRAS speed estimator, rotor flux, sensorless vector control**

I. INTRODUCTION

Induction motor drives are one of the most popular electrical machines used in different industrial branches because of its simple construction. Induction motor can be controlled using vector control, which provides high accuracy and dynamics control.

Nowadays the sensorless control of induction motor is extensively investigated, because elimination of speed sensor provides the advantages of reduced cost and size of electrical drive, higher reliability and less maintenance requirements. Various methods have been proposed for the speed estimation of induction motor, such as open-loop estimator, estimator using saturation of the third-harmonic voltage, model reference adaptive systems (MRAS) with different kind of observers, adaptive observers like Luenberger observer, Kalman filter or sliding mode observer, signal injection techniques and estimators using artificial intelligence.

MRAS schemes offer simpler implementation and require less computational effort compared to other methods based on mathematical model, hence are the most popular strategies used for sensorless control.

II. RF MRAS SPEED AND ROTOR FLUX ESTIMATOR

A. Mathematical model of induction motor

The mathematical model of an induction motor drive in a stationary frame (α, β) can be described by equations as follows:

The stator and rotor voltage equations:

$$\boldsymbol{u}_1^s = R_1 \boldsymbol{i}_1^s + \frac{d\boldsymbol{\Psi}_1^s}{dt} \tag{1}$$

$$0 = R_2 \boldsymbol{i}_2^s + \frac{d\boldsymbol{\Psi}_2^s}{dt} - j\omega_2 \boldsymbol{\Psi}_2^s \tag{2}$$

The stator and rotor flux linkages in the (α, β) reference frame:

$$\boldsymbol{\Psi}_1^s = L_1 \boldsymbol{i}_1^s + L_h \boldsymbol{i}_2^s \tag{3}$$

$$\boldsymbol{\Psi}_2^s = L_2 \boldsymbol{i}_2^s + L_h \boldsymbol{i}_1^s \tag{4}$$

where R_1 and R_2 are stator and rotor resistance respectively, i_1 and i_2 are stator and rotor current, L_1, L_2 and L_h are stator, rotor and mutual inductance, ω_2 is the rotor speed.

The electromagnetic torque and the rotor speed are related by the dynamic equation:

$$m_e - m_z = \frac{J}{p}\frac{d\omega_2}{dt} \tag{5}$$

where J is the inertia, p is number of pole-pairs and m_e is electromagnetic torque, which can be expressed:

$$m_e = \frac{3p}{2}\frac{L_h}{L_2}\left(\Psi_{2\alpha} i_{1\beta} - \Psi_{2\beta} i_{1\alpha}\right) \tag{6}$$

B. Rotor flux MRAS speed estimator

Usually the classical rotor flux MRAS (RF MRAS) speed estimator structure consists of a reference model, an adaptive model and an adaptation mechanism which generates the estimated speed. The reference model, called voltage model (VM), can be obtained from stator voltage equation and its output are the reference values of the rotor flux components. The reference rotor flux components can be obtained from the reference model as follows [1]:

$$s\Psi_{2\alpha} = \frac{L_2}{L_h}\left(u_{1\alpha} - R_1 i_{1\alpha} - \sigma L_1 s i_{1\alpha}\right) \tag{7}$$

$$s\Psi_{2\beta} = \frac{L_2}{L_h}\left(u_{1\beta} - R_1 i_{1\beta} - \sigma L_1 s i_{1\beta}\right) \tag{8}$$

where σ is the leakage coefficient and is defined:

$$\sigma = 1 - \frac{L_h^2}{L_1 L_2} \tag{9}$$

The adaptive model, called current model, can be obtained from the rotor voltage equation and its output represents the estimated value of the rotor flux components. Adaptive model inputs are stator current components expressed in (α, β) reference frame and estimated rotor speed $\hat{\omega}_2$. The identified

rotor flux components can be obtained by equations [1]:

$$s\hat{\Psi}_{2\alpha} = \frac{L_h}{T_2}i_{1\alpha} - \frac{1}{T_2}\hat{\Psi}_{2\alpha} - \hat{\omega}_2\hat{\Psi}_{2\beta} \qquad (10)$$

$$s\hat{\Psi}_{2\beta} = \frac{L_h}{T_2}i_{1\beta} - \frac{1}{T_2}\hat{\Psi}_{2\beta} + \hat{\omega}_2\hat{\Psi}_{2\alpha} \qquad (11)$$

where T_2 is rotor time konstant.

Design of the adaptation mechanism is an important part of the RF MRAS speed estimator and is based on Popov´s hyperstability criterion [5]. The adaptation mechanism minimizes the error between the reference and estimated rotor fluxes. First, it is necessary to define a speed tuning signal ε_ω, which is minimized by PI controller. The PI controller output is the estimated rotor speed which is fed back to the adaptive model. The speed tuning signal and the estimated speed can be expressed as [1]:

$$\varepsilon_\omega = \Psi_{2\beta}\hat{\Psi}_{2\alpha} - \Psi_{2\alpha}\hat{\Psi}_{2\beta} \qquad (12)$$

$$\hat{\omega}_2 = (k_p + \frac{k_i}{s})\varepsilon_\omega \qquad (13)$$
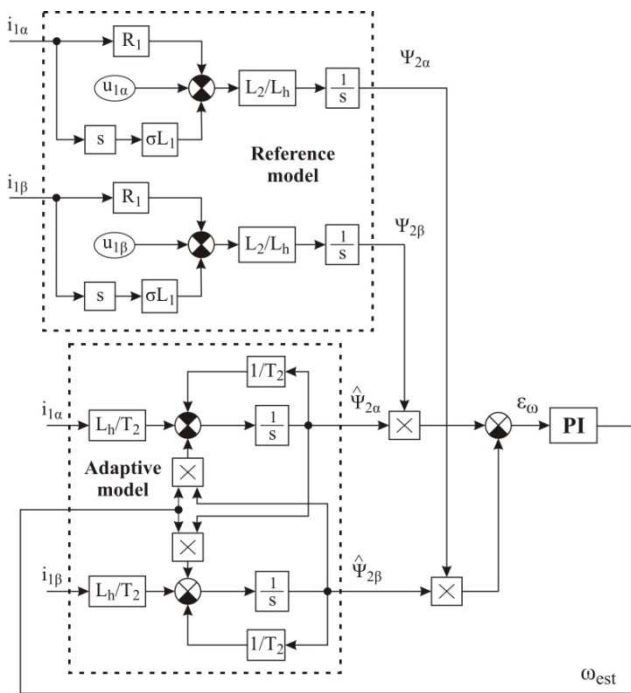
Fig. 1   Block diagram of RF MRAS estimator

III.   SIMULATION RESULTS

The RF MRAS estimator properties were verified in the structure of indirect vector control. The induction machine parameters are given in Table 1.

TABLE I

Parameter	Value [unit]	Parameter	Value [unit]
P_n	750 W	J	0,0054 kgm^2
U_n	220/380 V	R_1	9,978 Ω
I_n	3,8/2,2 A	$R_2´$	9,933 Ω
n_n	1380 rpm	L_h	0,4494 H
p	2	$L_{1\sigma} = L_{2\sigma}$	0,0368 H
S	0,08	$L_1 = L_2$	0,4862 H

The simulations have been carried out using MATLAB/Simulink platform and overall block diagram is shown in Fig. 2.
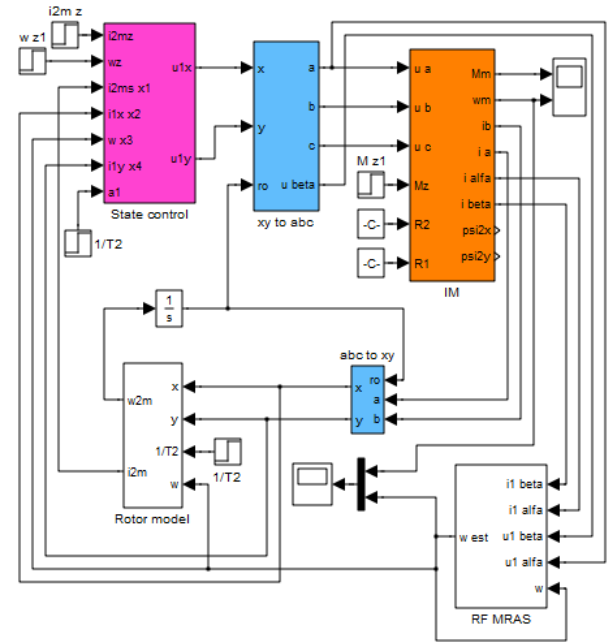
Fig. 2   Overall block diagram of sensorless vector control

Fig. 3 shows the speed response of the induction motor and RF MRAS estimator during speed reversion. Corresponding rotor flux waveforms (only beta components of reference and adaptive model) are shown in Fig. 4.
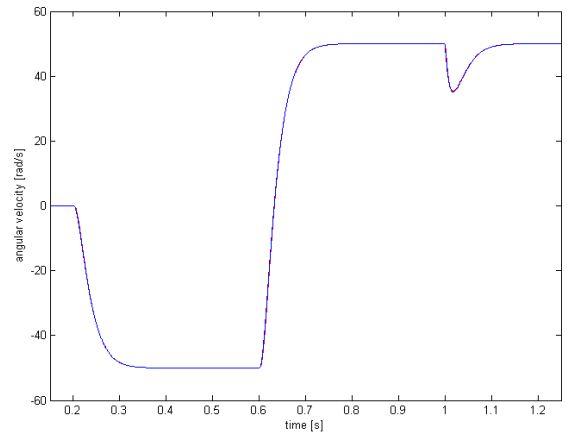
Fig. 3   Speed response during speed reversion from -50 to 50 rad/s
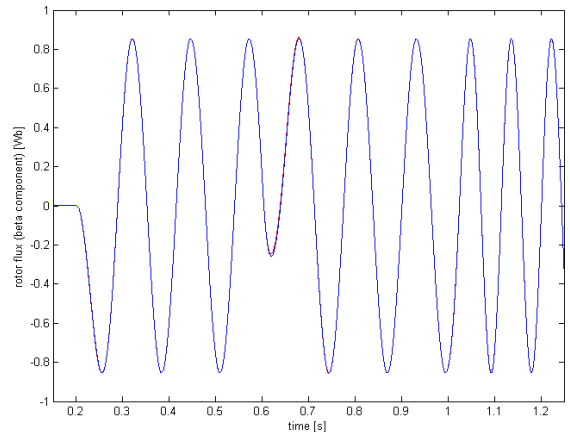
Fig. 4   Rotor flux waveforms of the reference and adaptive model

Fig. 6 Effect of incorrect setting of R_1 value to the speed response and rotor flux components (ω_{ref} = 80 rad/s, R_1 is increased by 10 %)

Fig. 5 Effect of incorrect setting of R_1 value to the speed response and rotor flux components (ω_{ref} = 30 rad/s, R_1 is increased by 10 %)

Fig. 5 shows effect of incorrect setting of R_1 value to the speed and beta rotor flux components of the reference and adaptive model RF MRAS estimator. From the figure is seen that the RF MRAS estimator is unstable. If the speed is increased to 80 rad/s (Fig. 6), the speed estimator can work properly, but due to an incorrect value of R_1 the speed waveforms oscillate.

IV. CONCLUSION

At the present time big effort of researchers is focused to the sensorless control of electrical drives, because elimination of speed sensor has several advantages. In this paper was discussed MRAS speed estimator based on rotor flux and it was implemented in the indirect vector control structure of induction motor. This type of speed estimator has very simple structure, but the other hand is highly sensitive to machine parameter variations, since is based on mathematical model. Particularly stator resistance variations present big problem at low speed range, where the estimator is unstable. Therefore it is necessary monitoring changes of stator resistance.

REFERENCES

[1] S. M. Gadoue, D. Giaouris, J. W. Finch, "Performance evaluation of a sensorless induction motor drive at very low and zero speed using a MRAS speed observer," *Industrial and Information Systems*, 2008. ICIIS 2008.IEEE Region 10 and the Third international Conference on, vol., no., pp. 1-6, 8-10 Dec. 2008.
[2] A. R. Haron, N. R. N. Idris, "Simulation of MRAS-based speed sensorless estimation of induction motor drives using MATLAB/SIMULINK" . *Power and Energy Conference, 2006. PECon 06. IEEE International,* vol., no., pp. 411-415, 28-29 Nov. 2006
[3] P. Vas, "Sensorless vector and direct torque control," New York: Oxford University press, 1998.
[4] J. W. Finch, D. Giaouris, "Controlled AC electrical drives," IEEE Transactions on Industrial Electronics, vol. 55, pp. 1-11, 2008. unpublished.
[5] C. Schauder, "Adaptive speed identification for vector control of induction motors without rotational transducers," *IEEE Transactions on Industry Applications,* vol. 28, pp. 1054-1061, 1992.
[6] S. Tamai, H. Sugimoto, M. Yano, "Speed-sensorless vector control of induction motor with model reference adaptive system," Conf. Record of the 1985 *IEEE-IAS Annual Meeting,* pp. 613-620, 1985.
[7] H. Yushui, L. Dan, "Realization of sensorless vector control system based on MRAS with DSP," Control Conference, 2008. CCC 2008. 27 th Chinese, vol., no., pp. 691-694, 16-18 July 2008.

14

Video Driver Assistance System Using Optical Correlator

[1]Tomáš HARASTHY, [2]Anna KOLESAROVÁ KAŽIMÍROVÁ

[1,2] Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]tomas.harasthy@tuke.sk, [2]anna.kolesarova@tuke.sk

Abstract — **Paper presents experiments and results with traffic sign recognition system based on optical correlator. The traffic scene is captured with color camera and key frame is extracted as input for the system. The system consists of two main blocks, Preprocessing and Optical Corelator. The Cambridge Correlator was used. The ROI is defined and chosen in preprocessing block. Preprocessed ROI go to Optical Correlator a then is compared with reference images (traffic signs). Several experiments have been done with this system. Results and conclusion are discussed.**

Keywords— **Optical correlator, Traffic Sign, color segmentation, Optical Fourier Transform, Joint Power Spectrum**

I. INTRODUCTION

Video-based object recognition has many applications in different fields [1-4]. Each of recognition systems has its own specific requirements for input data (image or video). Input data have to be preprocessed, to producing relevant features to next processing.

This paper is concentrated to Drive Assistance System based on Optical Correlator. Driver Assistance System helps the driver in his drive process and its most important task is driver and crew safety.

Goal of this paper is present experiments with Driver Assistance System based on Cambridge Correlator.

II. SYSTEM FOR TRAFFIC SIGN RECOGNITION

Traffic sign recognition is one of most important part of Drive Assistance System [1-4,7]. For traffic sign recognition is fast process time very important. In Fig. 1 is shown block diagram of proposed system to recognize traffic signs using Optical Correlator.
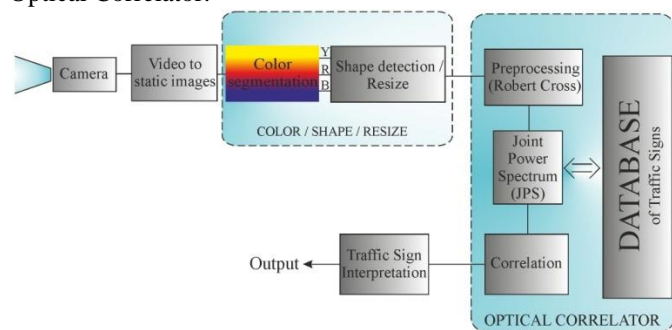
Fig. 1. Proposed Trafic Sign Recognition System.

Traffic Sign Recognition System (Fig. 1) consists of two main blocks, Preprocessing [7] and Optical Correlator.

Traffic scene is captured with the color camera. From video stream the key frame is extracted and then used in next processing. Next processing consists of converting key frame from RGB to HSV color space. HSV color space is represented by three components: Hue (H), Saturation (S) and Value (V).

Fig. 2. Color Filtering.

Fig. 3. Defining ROI
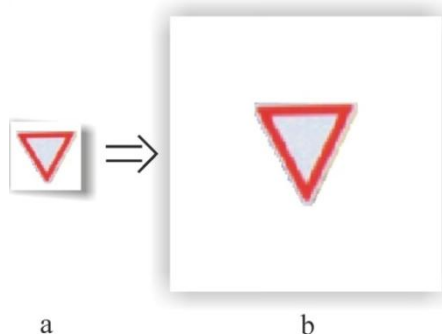
Fig. 4.   Resize input image for Optical Correlator

First step of preprocessing is exposition correction. In next step of preprocessing, the color filter is used. Color filters are designed to remove all irrelevant colors in input traffic scene. After color filtering, just relevant color is available. If the "RED" filter is used, bitmap looks such us Fig. 2b and if the "BLUE" filter is used, bitmap looks such us Fig. 2c. For "Yellow" filter no relevant color was detected so the bitmap looks such us white image, without any features. Images after color filtering contains color clusters, which are irrelevant for detecting traffic signs. In next step, the irrelevant color clusters are deleted (Fig. 2d,e). After color filtering and irrelevant cluster removing, the potential Region of Interest (ROI) is shoved (Fig. 2d,e). Minimum and maxim size of ROI has to be defined to ignoring too big and too small objects. The ROI is defined at the input image. ROI crop input image to chosen traffic sign (Fig. 3). The background is removed from input image and then traffic sign without background is resized to correct size for input of Optical Correlator (Fig. 4).

III. EXPERIMENTS AND RESULTS

In our experiments the "Speed limit" traffic signs were used. Three traffic sign were chosen and tested. All speed limits traffic signs contains red circle, so in preprocessing the "RED" filter was used. Every red circle on traffic scene was defined as a ROI. After defining ROI, the background was removed. The red circle of traffic sign was defined as a background, because every speed limit traffic sign contains this circle and correlation is too high between every sign. Background was deleted first to red color, and then to black color. After this preprocessing, just black symbols inside the red circle was left. As input to correlator were used only numbers 40, 60 and 80. The Fig. 5 show example of the input to correlator for speed limit 40 (a), JPS (b), JPS preprocessed – threshold (c) and correlation (d).

Correltaion was found between chosen traffic sign, but some peaks were founded between 40-60 and 40-80 as well. Similarity between different numbers was much lower. All tested traffic signs contains number "0", so similarity was found between all tested images. Similarity between same signs, for example 40-40, was much higher than between 40-60, 40-80 or 60-80 (Fig. 6).

On the Fig. 6 is shown, that correlations occurs between all tested traffic signs, but correlation between same symbols (40 - 40) is the highest. On the Fig. 7 is shown correlation between tested signs and each correlation is coupled with 3D graph produced in MATLAB.
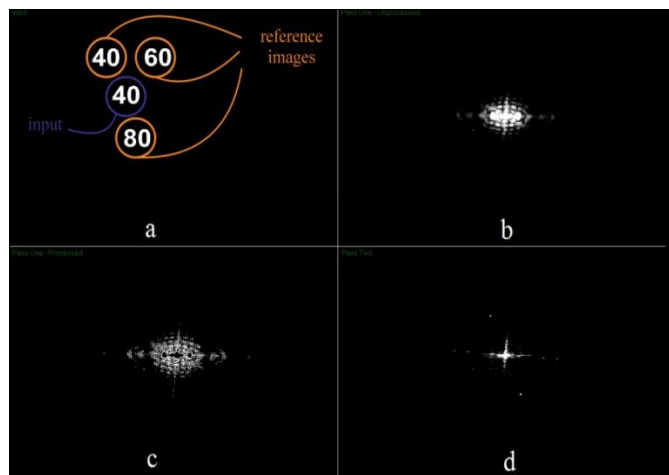
Fig. 5.   Correlation process
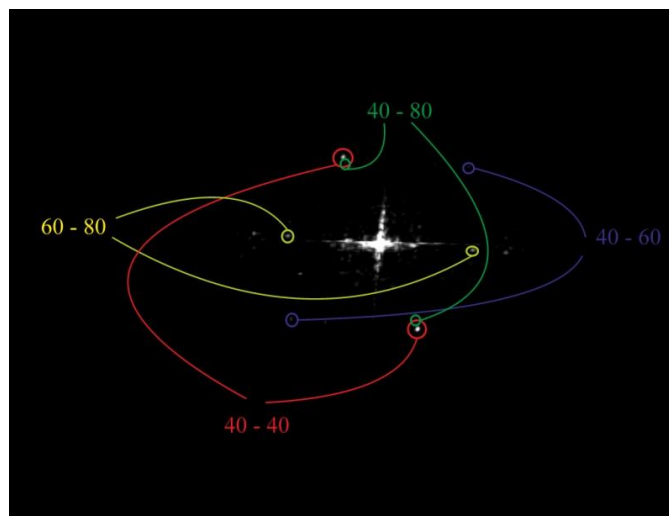
Fig. 6.   Correlations

On the Fig. 7a is shown "Speed Limit 40", Fig. 7b "Speed Limit 60" and Fig. 7c "Speed Limit 80".On these graphs is shown, that correlation between same images is much higher than peaks of other correlations founded in image.

IV. CONCLUSION

In this paper the Traffic Signs Recognition system with Using Optical Correlator was presented. Traffic scene was captured with color camera and then the key frame was extracted. The key frame was preprocessed and converted to HSV color space. "RED", "BLUE" and "YELLOW" bitmap was extracted.

In experiments, the "Speed Limit" traffic signs were used. The preprocessing was concentrated to red circle. Just symbols inside red circle were relevant. These symbols were compared with Optical Correlator.

In all tested "Speed Limit" traffic signs were correct signs recognized. Correlation between same signs (symbols) was much higher than correlations between others signs, so correct traffic signs could be recognized.

In future we want to focus preprocessing input information and use preprocessed video as input for Optical Correlator. Another target is using these system to recognition all Traffic Signs, not just "Speed Limit". These can be done with better preprocessing and using better "Reference image" in Optical Correlator.
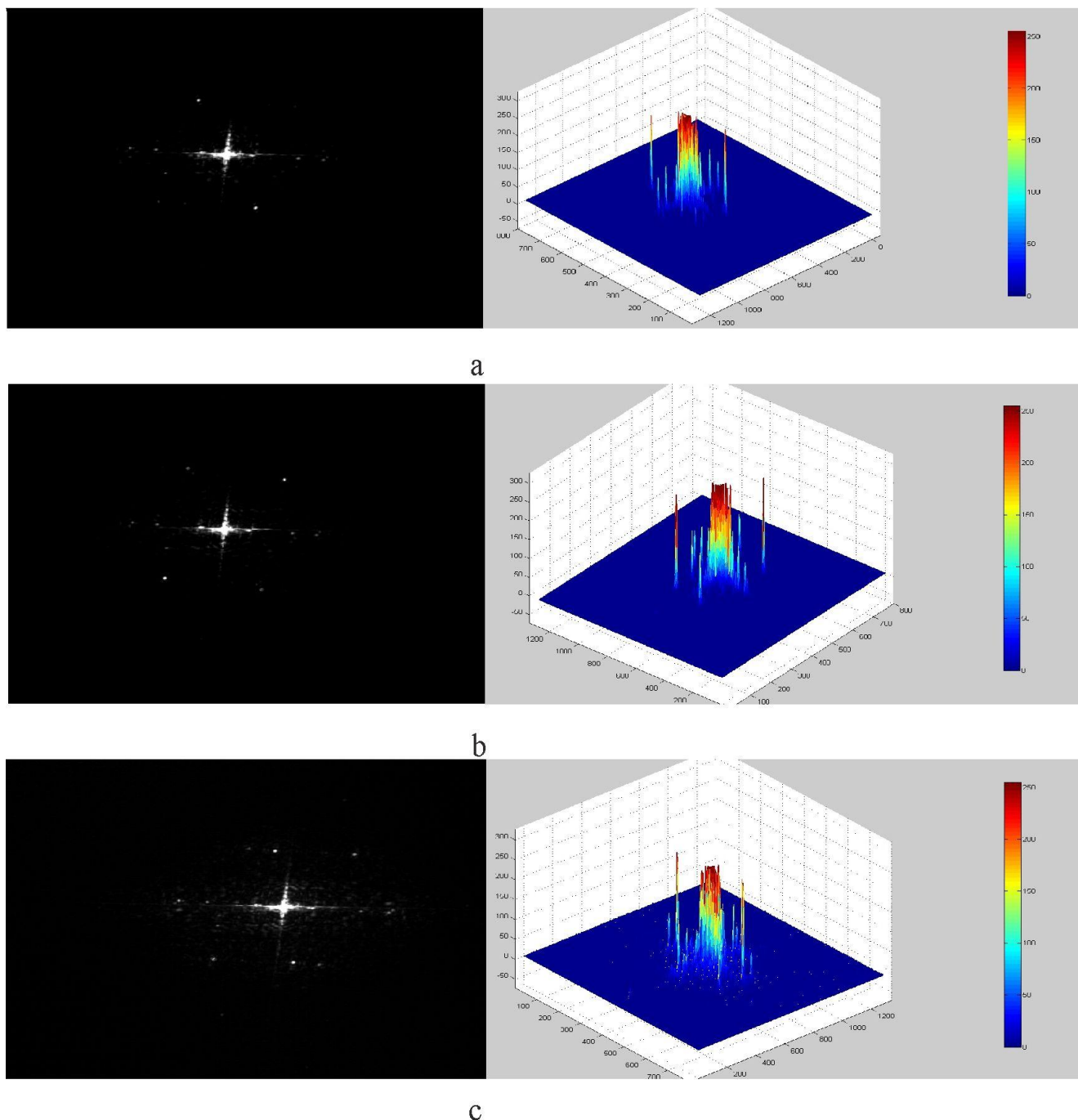
a

b

c

Fig. 7. Correlation peaks

REFERENCES

[1] F. P. Paulo, L. P. Correia, "Automatic detection and classification of traffic signs," *Image Analysis for Multimedia Interactive Services,* WIAMIS, 6-8 June 2007.

[2] A. Laika, W. Stechele, "A review of different object recognition methods for the application in driver assistance systems, " in *Image Analysis for Multimedia Interactive Services,* WIAMIS, 6-8 June 2007.

[3] J. Turán, M. Fifik, Ľ. Ovseník and J. Turán Jr., "Transform based system for traffic sign recognition," *15th International Conference on Systems, Signals & Image Processing,* IWSSIP, Bratislava, Slovakia, 2008.

[4] A. Broggi, P. Cerri, P. Medici, P. P. Porta, "Real Time Road Signs Recognition", *Intelligent Vehicles.*

[5] Cambridge Correlators, http://www.cambridgecorrelators.com, (february 2011).

[6] Cambridge Correlators, "Fourier Optics Experimenter",*User Guide.*

[7] M. Fifik, J. Turán, Ľ.Ovseník, K.Fazekas, "Experiments with a Transform based Traffic Sign Recognition System," *Proc. of 17th International Conference on Systems, Signals and Image Processing,* IWSSIP2010, Rio de Janeiro, Brazil, June 17-19, 2010, 227-230.

Video Surveillance System Using Optical Correlator

[1]*Anna KAŽIMÍROVÁ KOLESÁROVÁ*, [2]*Tomáš HARASTHY*

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]anna.kolesarova@tuke.sk, [2]tomas.harasthy@tuke.sk

Abstract—**This paper is review of practical video surveillance systems. With the growing quantity of security video, it becomes vital that video surveillance systems be able to support security personnel in monitoring and tracking activities. The aim of the surveillance applications is to detect, track and classify targets. We will also describe a design of our video surveillance system with optical correlator.**

Keywords — *Video surveillance systems, surveillance applications, security video, security personnel, optical correlator.*

I. INTRODUCTION

With the growing quantity of security video, it becomes vital that video surveillance systems be able to support security personnel in monitoring and tracking activities. Video surveillance systems are common in many environments. Video surveillance has been a key component in ensuring security at airports, metros, casinos, banks and correctional institutions. More recently, government agencies, businesses, and even schools are turning toward video surveillance as a means to increase public security [5].

Several important research questions remain to be addressed before we can rely upon video surveillance as an effective tool for crime prevention, crime resolution, and crime protection [2].

Some example applications include intrusion detection, activity monitoring, and pedestrian counting. The capability of extracting moving objects from a video sequence is a fundamental and crucial problem of these vision systems [3].

II. OUR SYSTEM USING OPTICAL CORRELATOR

We designed system (see Fig. 1) that works follow s. We have video output from CCD camera. This video output is divided to video sequences that are input for process called pre-processing. To recognition moving objects on the background, head detection and luggage detection we used the tracker. *Tracker* contained following blocks*: Motion Detector, Head Detector, Shape Tracker* and *Region Tracker*. Tracking output is recognized in recognition block [4].

A Motion Detector detects moving pixels in the image. It models the background as an image with no people in it. Simply subtracting it pixel wise from of the current video image and thresholding the result yields the binary motion image. Regions (bounding boxes) with detected moving blobs are then extracted and written out as the output from this module. Main features of a Motion Detector are:

- Simple background image subtraction,
- Image filtering (spatial median filter, dilation) depending on available CPU time,
- Temporal inclusion of static objects into the background,
- Background modelling using a speed-optimized median filter,
- Static regions incorporated into background (multi-layer background).

A Head Detector makes rapid guesses of head positions in all detected moving regions. Main features of a Head Detector are:

- Works in binary motion image,
- Looks for peaks in detected moving regions,
- Vertical pixel histogram with low-pass filter,
- Optimized for speed not accuracy.

A Shape Tracker uses a deformable model for the 2D outline shape of a walking pedestrian to detect and track people. The initialization of contour shapes is done from the output by the Region Tracker and the Head Detector. Main features of a Shape Tracker are:

Fig. 1. System block diagram.

- Local edge search for shape fitting,
- Initializing of shape from Region Tracker, Head Detector and own predictions,
- Occlusion reasoning.

A Region Tracker tracks these moving regions over time. This includes region splitting and merging using predictions from the previous frame. Main features of a Region Tracker are:

- Region splitting and merging using predictions,
- Adjust bounding box from Shape Tracker results,
- Identify static regions for background integration.

Recognition block contained two blocks: *Classificator* and *Personal Recognition*. Data from recognition output are compared with data from *Image Memory*.

Image memory is database of static images of human faces, that have guarded enter to this room (employees faces).

Learning is a process of personal identities creation.

A device for comparing two signals by utilizing the Fourier transforming properties of a lens is *optical correlator*. It is commonly used in optics for target identification and tracking. The correlator has an input signal which is multiplied by some filter in the Fourier domain. An example filter is the matched filter which uses the cross correlation of the two signals. An optical correlator automatically identifies or recognizes the contents of an image by combining an incoming image with a reference image, and the degree of correlation after combining the images determining the intensity of an output light beam [1].

III. EXPERIMENT AND RESULTS

A new method for a robust and efficient analysis of video sequences is presented; it allows the extraction of foreground objects and the classification of static foreground regions as abandoned or removed objects.

On Fig. 2 is shown how to works our face detection with recognition system. The face detection and recognition system utilizes a video camera and a PC with optical correlator to perform the person identification task. It should:

1. Detect faces from the still images (or video sequences). There is no restriction to the background and the number of faces.

2. Handle the varying conditions in a consumer environment.

3. Recognize faces by comparing the captures face images to a database of known faces. A decision is then made about the identity of the person (either one of the faces in the database, or not belong to the database at all).

4. Enable that the face detection and recognition stage can take place at different locations [4].

IV. CONCLUSION

Surveillance systems significantly contribute to situation control. Such systems transform video surveillance from a data acquisition tool to information and intelligence acquisition systems. Real-time video analysis provides surveillance systems with the ability to react to an activity in
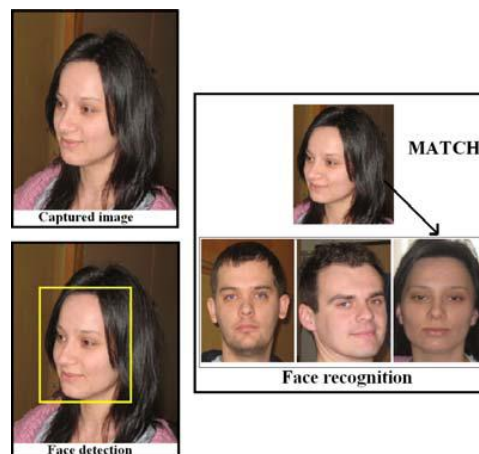
Fig. 2. Face detection a recognition system.

real time, thus acquiring relevant information at much higher resolution [1]. The long-term operation of such systems provides the ability to analyze information in a spatial-temporal context.

The problem of remote surveillance has received growing attention in recent years, especially in the context of public infrastructure monitoring for transport applications, safety of quality control in industrial applications, and improved public security. The development of a surveillance system requires multidisciplinary expertise, including knowledge of signal and image processing, computer vision, communications and networking pattern recognition and sensor development and fusion [1].

Our system is preventing before entering forbidden person and leaving the suspicious luggage into the guarded room. In this luggage or package could be bomb, gun, drugs, etc. On the other side, big task is checking if some person steals the luggage, package or the other things.

Our system could increase security employees and the other people in schools, stations, airports, etc.

REFERENCES

[1] BHARGAVA, M. – CHEN, CH. - RYOO, M. S. - AGGARWA, J. K.: „*Detection of object abandonment using temporal logic*", Springer Berlin, pp. 271-281, January 2009.

[2] DIMITRIJEVIČ, Z. - WU, G. – CHANG, E.: *"A Multi-Sensor Fusion and Mining System"*, Proceeding of the 2nd Usenix FAST, March 2003.

[3] OLIVEIRA, R. J. - RIBERIO, P. C. - MARQUES, J. S. – LEMOS, J. M.: *"A Video System for Urban Surveillance: Function Integration and Evaluation"*, VIAMIS, 2004.

[4] OVSENÍK, Ľ. – KAŽIMÍROVÁ KOLESÁROVÁ, A. – TURÁN, J.: „*Video Surveillance Systems*", In: Acta Electrotechnica et Informatica, Košice : FEI TU,Vol. 10, No. 4, 2010, pp. 46-53.

[5] RANGASWAMI, R. - DIMITRIJEVIC, Z. - KAKLIGIAN, K. - CHANG, E. – WANG, Y. F.: *"The SfinX Video surveillance system"*, 2003.

Dynamic Analysis of AC Machines Properties Using Virtual Models in the MATLAB GUI

Tibor Balogh

Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

tiborb@gmail.com

Abstract—**The paper describes advantages and building virtual models in Graphical User Interface (GUI) MATLAB environment. The virtual models of the BLDC motor and induction motor were developed by applying mathematical equations. We have used a BLDC model with trapezoidal waveforms of back-EMF voltages and induction motor in XY coordinate system. Based on analysis of time responses a comparison study of results of both motors is presented in GUI MATLAB environment.**

Keywords—**Induction motor, brushless DC motor, modeling, simulation**

I. INTRODUCTION

Virtual models have undergone rapid progress in recent years and have been widely accepted not only for commercial product development, but for eLearning and lectures, too. Product design and manufacturing organizations are moving from the traditional, multiple and serial design-build-test cycle approach to simulation-led problem solving and performance validation using CAE, CAD or MATLAB/Simulink tools [1].

Verification of the theoretical knowledge by real machines is risky, slow and costly. Virtual models of the machines can overcome this difficulty.

II. CONSTRUCTION AND OPERATING PRINCIPLE

The BLDC motor is also referred to as an electronically commuted motor. There are no brushes on the rotor and the commutation is performed electronically at certain rotor positions. The stator magnetic circuit is usually made from magnetic steel sheets. The stator phase windings are inserted in the slots (a distributed winding), or can be wound as one coil on the magnetic pole. The magnetization of the permanent magnets and their displacement on the rotor are chosen in such a way that the back-EMF shape is trapezoidal. This allows the three-phase voltage system, with a rectangular shape, to be used to create a rotational field with low torque ripples. In this respect, the BLDC motor is equivalent to an inverted DC commutator motor, in that the magnet rotates while the conductors remain stationary. In the DC commutator motor, the current polarity is reversed by the commutator and the brushes, but in the brushless DC motor, the polarity reversal is performed by semiconductor switches which are to be switched in synchronization with the rotor position. Besides the higher reliability, the missing

commutator brings another advantage. The commutator is also a limiting factor in the maximal speed of the DC motor. Therefore the BLDC motor can be employed in applications requiring high speed.

The simple motor model of a BLDC motor consisting of a 3-phase power stage and a brushless DC motor is shown in Fig. 1.
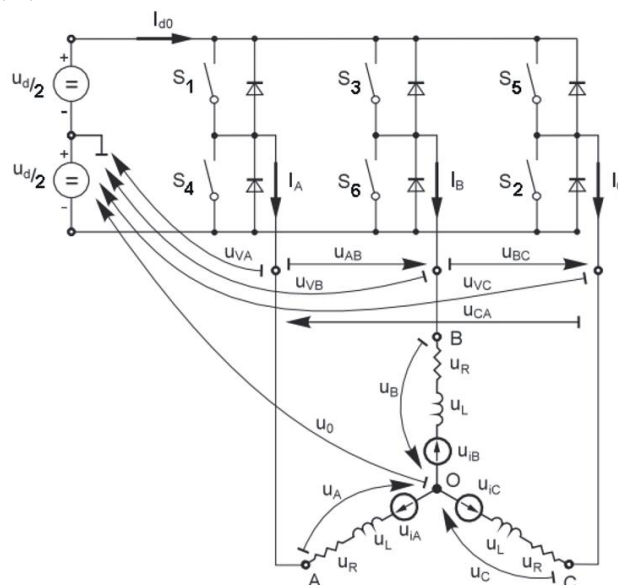
Fig. 1. BLDC motor model [2]

III. MATHEMATICAL MODEL OF THE MOTOR

Modeling of a BLDC motor can be developed in the similar manner as a three-phase synchronous machine. Since there is a permanent magnet mounted on the rotor, some dynamic characteristics are different. Flux linkage from the rotor depends upon the magnet material. Therefore, saturation of magnetic flux linkage is typical for this kind of motors. As any typical three-phase motors, one structure of the BLDC motor is fed by a three-phase voltage source. The source is not necessarily to be sinusoidal. Square wave or other wave-shape can be applied as long as the peak voltage does not exceed the maximum voltage limit of the motor. Similarly, the model of the armature winding for the BLDC motor is expressed as follows:

$$v_a = Ri_a + L\frac{di_a}{dt} + e_a \tag{1}$$

$$v_b = Ri_b + L\frac{di_b}{dt} + e_b \qquad (2)$$

$$v_c = Ri_c + L\frac{di_c}{dt} + e_c \qquad (3)$$

where
L is armature self-inductance [H],
R - armature resistance [Ω],
v_a, v_b, v_c - terminal phase voltages [V],
i_a, i_b, i_c - motor input currents [A],
and e_a, e_b, e_c - motor back-EMFs [V].

In the 3-phase BLDC motor, the back-EMF is related to a function of rotor position and the back-EMF of each phase has 120º phase angle difference so equation of each phase should be as follows:

$$e_a = K_w f(\theta_e)\omega \qquad (4)$$

$$e_b = K_w f(\theta_e - 2\pi/3)\omega \qquad (5)$$

$$e_c = K_w f(\theta_e + 2\pi/3)\omega \qquad (6)$$

where
K_w is back EMF constant of one phase [V/rad.s^{-1}],
θ_e - electrical rotor angle [° el.],
ω - rotor speed [rad.s^{-1}].

The electrical rotor angle is equal to the mechanical rotor angle multiplied by the number of pole pairs:

$$\theta_e = \frac{p}{2}\theta_m \qquad (7)$$

Where θ_m presents mechanical rotor angle [rad].

Total output torque can be represented as summation of that of torque of each phase. The next equation represents the total output torque:

$$T_e = \frac{e_A i_A + e_B i_B + e_C i_C}{\omega} \qquad (8)$$

where
T_e is total output torque [Nm],
p - number of pole pairs.

The equation of mechanical part is represented as follows:

$$T_e - T_l = J\frac{d\omega}{dt} + b\omega \qquad (9)$$

where
T_l is load torque [Nm],
J - rotor inertia [kgm^2],
B - friction constant [Nms.rad^{-1}].

The detailed mathematical model of the induction motor is generally known and due to this fact it not mentioned here (also saving the allocated space for the contribution).

IV. SIMULINK MODEL OF THE MACHINES

Fig. 2 shows the BLDC drive motor SIMULINK model. The model was developed in the rotor reference frame of the BLDC motor using equations shown above [3].
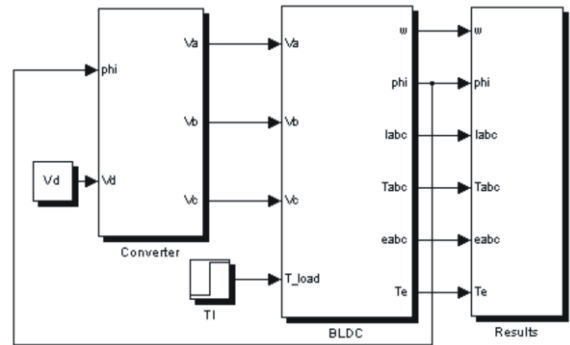
Fig.2. SIMULINK model of the BLDC drive

Fig. 3. shows the detail of the BLDC motor block.
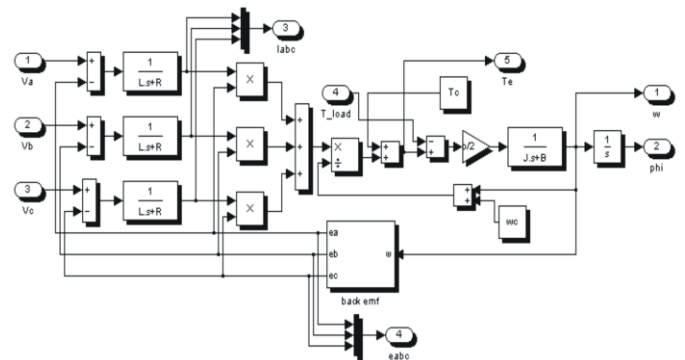
Fig. 3. SIMULINK model of the BLDC motor block

Fig. 4. shows the detail of the induction motor block.
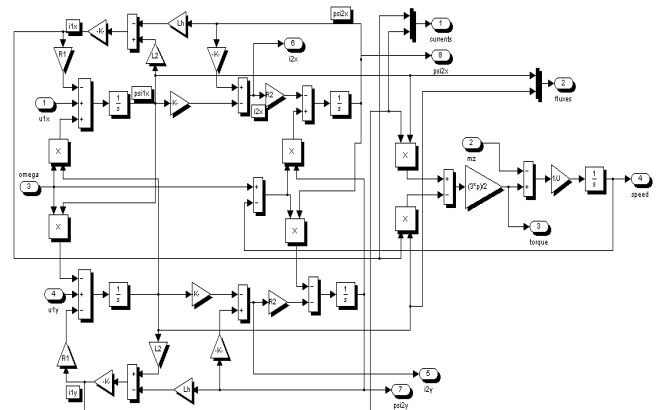
Fig. 4. SIMULINK model of the induction motor block

Fig. 5 shows results of simulation of a BLDC motor with the following parameters: V_d = 30V, R = 4,98Ω, L = 5,05 mH, p = 4, J = 15,17.10^{-6} kgm^2, k_w = 56,23.10^{-3} V/rad.s^{-1}, T_l = 0,1 Nm. The load time is 0,15 s.

The time courses in Fig. 5 belong to the BLDC motor with trapezoidal back-EMF. There is a very good correlation with the responses of the BLDC motor models known in the references [3], [4].
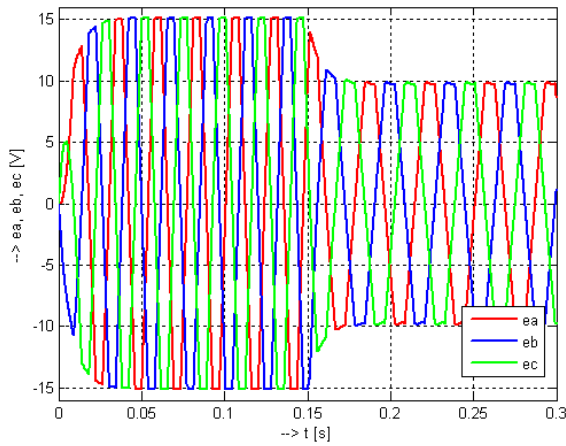
Fig. 5. Time courses of trapezoidal back-EMF

V. Graphical User Interface in Matlab

A graphical user interface provides the user with a familiar environment in which to work. A graphical user interface (GUI) presents a pictorial interface to the program. A good GUI development environment can make programs easier to use by providing them with a consistent appearance and with intuitive controls like pushbuttons, list boxes, sliders, menus, and so forth, all of which are already familiar to the user, so that he or she can concentrate on using the application rather than on the mechanics involved in doing things. The GUI itself should behave in an understandable and predictable manner, so that a user knows what to expect when he or she performs an action. For example, when a mouse click occurs on a pushbutton, the GUI should initiate the action described on the label of the button.

However, GUIs are harder for the programmer because a GUI-based program must be prepared for mouse clicks (or possibly keyboard input) for any GUI element at any time. If GUIs were not used, people would have to work from the command line interface, which can be extremely difficult and frustrating [5].

The three principal elements required to create a MATLAB Graphical User Interface are:
- Components
- Figures
- Callbacks

Development of GUI

In GUI program, we can simulate the equations that describe not only mechanical but also electrical transients. First, we have to open a MATLAB and type command *guide* in the command window. Then we should see the blank GUI screen and we can start with the virtual model building.

Fig. 6 is shown GUI before the final version with settings table for the pushbutton simulation.
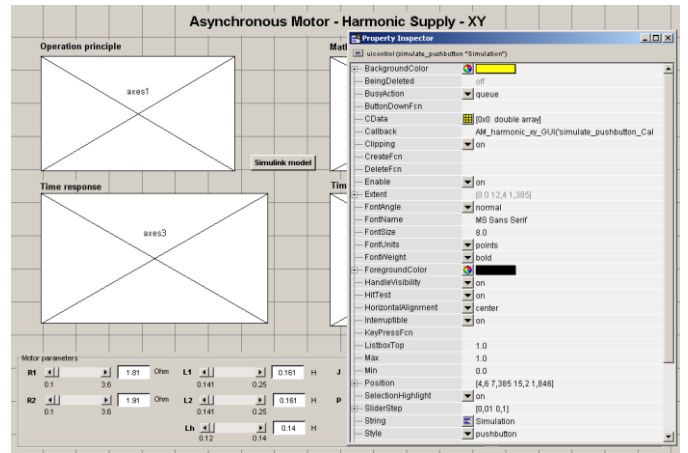
Fig. 6. GUI settings table for the pushbutton simulation

Fig. 7 is shown GUI for modeling and simulation of the BLDC motor. By blue pushbutton e_a, e_b, e_c we can see time courses of trapezoidal back-EMF in three-phase stator windings of the BLDC motor. Fig. 7 is shown mentioned back-EMF voltages before and after the load mount in time 0,15 s.
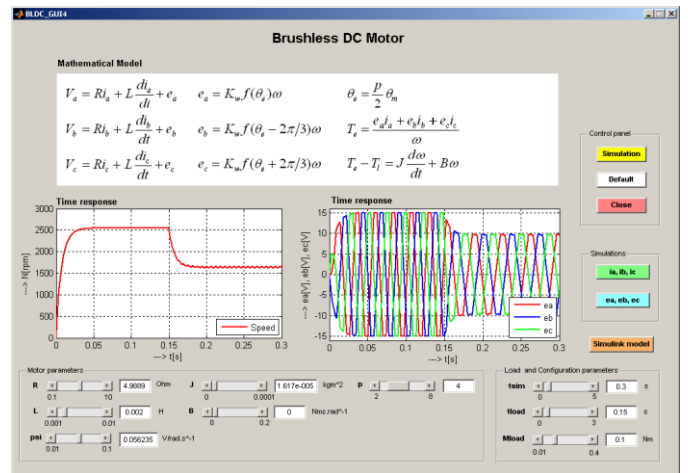
Fig. 7. Time courses of the BLDC motor in Matlab GUI

In the bottom of GUI simulation model are sliders to change the BLDC motor parameters. After changing the motor parameters we can monitor the behavior of the motor and show time courses of parameters that we want to monitor and analyze.

Students by GUI simulation can better understand the theoretical knowledge acquired from lectures into real devices. Via the GUI simulation we can monitor and analyze response time of physical quantities that are available.

Fig. 8 is shown a response time of the load variations of 0,1 Nm to 0,2 Nm. After changing we can analyze by the increase in the size of the load that speed of the motor and also back-EMF voltages in each three-phase in the stator winding were decreased.
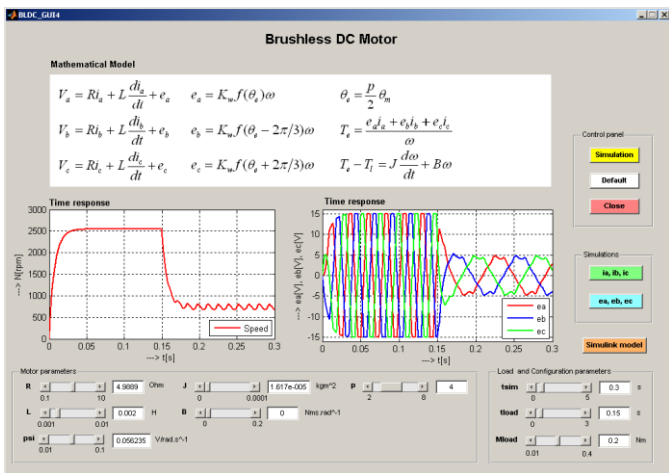
Fig. 8. Time courses of speed and back-EMF voltages

Fig. 9 is shown GUI for modeling and simulation of the induction motor. By green pushbutton i1x, i1y we can see time courses of currents in three-phase rotor windings of the induction motor. Fig. 9 is shown mentioned currents before and after the load mount in time 1,5 s.
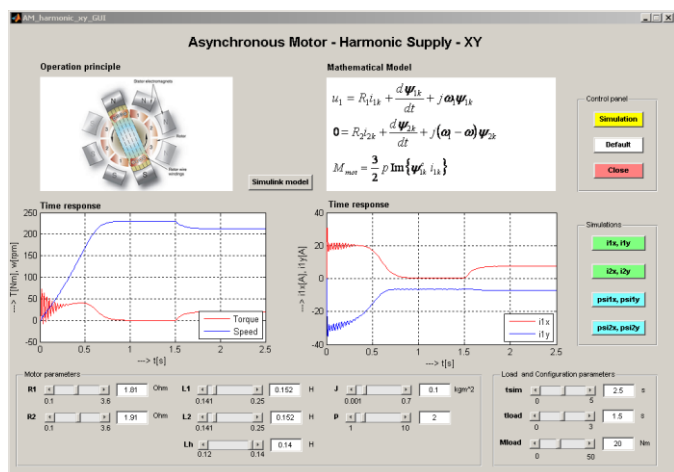
Fig. 9. Time courses of the induction motor in Matlab GUI

Fig. 10 is shown a response time of the rotor inertia variations of 0,1 kgm^2 to 0,191 kgm^2. After changing we can analyze by the increase in the size of the rotor inertia that the dynamic response of the speed of the motor was decreased.
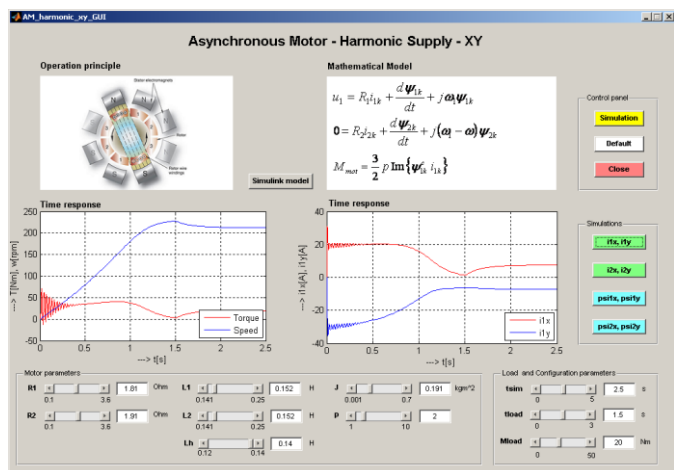
Fig. 10. Time courses of speed after higher value of the rotor inertia

VI. CONCLUSION

Presented virtual models are used in teaching in the field of electrical machines, drives and actuators. Students will get experiences in the analysis of virtual models before they verify properties of the machines measured on real devices in the laboratory.

The modeling procedure presented in this paper helps in simulation of the BLDC motor and induction motor. The performance evaluation results show that, such a modeling is very useful in studying the drive system before taking up the dedicated controller design, accounting the relevant dynamic parameters of the motor.

An important aspect in understanding the physical process of real devices is the acquisition of theoretical knowledge and simulation results. MATLAB GUI is tool that will help us to do.

REFERENCES

[1] T. J. Milburn, "The New Product Development Paradigm Led by Simulation and Testing", *SAE Technical Paper Series 2004-01-2667*.

[2] Pavel Grasblum, "Sensorless BLDC motor control using an 8-bit MCU", *Freescale*.

[3] S. Baldursson, "BLDC Motor Modelling and Control – A Matlab/Simulink Implementation", *Master Thesis, May, 2005*.

[4] W. Hong, W. Lee and B. K. Lee, "Dynamic Simulation of Brushless DC motor Drives Considering Phase Commutation for Automotive Applications", *IEEE International Electric Machines & Drives Conference, 2007. IEMDC 2007, pp.1377-1383, 3-5 May 2007*.

[5] Abdulla Ismail Abdulla, "Introduction to Graphical User Interface (GUI) MATLAB 6.5", *IEEE UAE Section*.

[6] Y.S. Jeon, H.S.Mok, G.H. Choe, D.K. Kim and J.S. Ryu, "A New Simulation Model of BLDC Motor with Real Back EMF waveforms", *IEEE CNF. On Computers in Power Electronics, 2000. COMPEL 2000. pp. 217 – 220, July 2000*.

[7] H. Kuchynkova, V. Hajek, "New Trend in Education – Virtual Model of Electrical Machines", *Zeszyty Problemowe – Maszyny Elektryczne Nr. 83/2009*.

[8] Ernest, Emil Sztylka, Rafal Ufnalski, Bartlomiej Koczara, Wlodzimierz, "Methods in Teaching Modern AC Drives: Inverter-fed Motor System with Internet-based Remote Control Panel", *12th International Power Electronics and Motion Control Conference, 2006. EPE-PEMC 2006. Portorož, Slovenia*

[9] V. Fedak, T. Balogh, P. Bauer, Š. Juško, "Virtual and Remote Experimentation in Motion Control", *11th International Conference on Mechatronics, "Mechatronika 2008", June 4-6, 2008, Trenčianske Teplice, Slovakia*.

Section:
Informatics & Telecommunications

Alarms in Supervisory Control

Roman MIHAĽ, Miloš PAVLÍK, Peter KARCH

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

roman.mihal@tuke.sk, milos.pavlik@tuke.sk, peter.karch@tuke.sk

Abstract—**Supervisory control represents a base for development of complete control system at present. The main elements of supervisory control applications are data acquisition from technological process level, process visualization with the control function and alarm situations indication. This article presents concept of alarms in supervisory control application using software products from Wonderware Corp.**

Keywords—**supervisory control, alarms, visualization.**

I. INTRODUCTION

Supervisory control system represents an inherent part of information and control system of industrial plant processes. At present, supervisory control applications even allocates big share of factory software investments.

In industrial sphere the issue of supervisory control is also known by the acronym SCADA/HMI (Supervisory Control And Data Acquisition / Human – Machine Interface) and presents central unifying position in the factory control system.

Alarm subsystem is the most important component of general realization of supervisory control application in conditions of real-time process control. It creates separate section in the visualization screens with basic features like signaling or acknowledgement of alarm states.

II. SUPERVISORY CONTROL AND ALARMS

A. Supervisory control

Supervisory control [1][2][3] represents a system, that provides monitoring and control of geographically large technological processes. Within this system there are included all the equipment and functionality for acquisition, processing, transmission and visualization of necessary information about controlled processes.

Supervisory control system in an industrial plant operates these primary functionalities:

- data acquisition from technological level of controlled system,
- visualization – alarms, trends, data and process flow presentation,
- technological process control,
- communication with machinery level and other subsystems at a higher control level,
- data archiving – cooperation with database system.

The most pronounced component of SCADA/HMI application is visualization.

Visualization of processes [4] means application of theoretical, technical, software and communication issues within industrial plant to visualize defined objects concerning technological processes and their automatic control system with the aim to support real-time decision making and control.

Current visualization software tools offer full graphical user interface for development and deployment of applications. In many cases this interface has the object oriented representation possibility of system elements or it contains preformatted libraries with graphical objects for different parts and sectors of industrial equipment.

B. Alarms

Alarm system as an important part of supervisory control is the most cardinal functionality of process monitoring and visualization. By means of different forms of alarm signaling [5] it provides human operators the possibility to observe information about overrunning admissible variable states, so subsequently they are able to perform the action, assess the situation and make a real-time response to changes in manufacturing.

Usually there are two basic alarm types that can be defined on process variables within the SCADA/HMI software:

1. alarms generated because of reaching the value limit of process variable,
2. change-rate alarms – caused by rapid change in value.

Alarm states, which are an inseparable component of supervisory control system, can be divided into following groups:

- process variables alarms,
- communication status alarms,
- diagnostic alarms of controlled process,
- SCADA/HMI system state alarms,
- alarms generated because of security violation of control system (Intrusion Detection Systems).

Whereas the alarm subsystem of HMI application is very important, it must be designed in good enough visible part of visualization screens. Visual differentiation of alarm states presents the most relevant task for development of the screens. Alarms signaling should be easily visual identified by the operator, so they can analyze alarm state and make changes in controlled system or acknowledge the alarms.

III. ALARM SYSTEM CONFIGURATION

In this section of article is described alarm configuration in ArchestrA IDE (Integrated Development Environment) as an environment of industrial software products from Wonderware Corp.

The best known product is InTouch that provides development and running of complete visualization applications with predefined graphic tools. A superstructure above InTouch makes product Wonderware Application Server, which introduces technology ArchestrA presented as unifying software architecture with products and services for effective design and deployment of automation and information application in industrial sphere.

ArchestrA technology interconnect functionality of InTouch and Application Server into development environment ArchestrA IDE where is used object-based representation for conveying process variables and complex of elements with predefined options.

Both InTouch and Application Server (ArchestrA) have their own engine tools for alarm management configuration. Fundamental difference between them is in alarm manager engine, in ArchestrA there is configuration of alarms performed on objects representing process variables (Fig. 1) rather than only on process variables in InTouch.
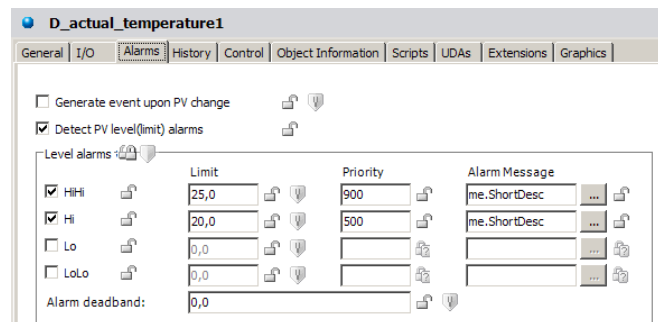
Fig. 1 Alarm with levels and priority definition on object representing process variable in ArchestrA IDE

The alarm agent marked as AlarmClient in ArchestrA IDE can be found in the section of graphical tools and symbols called Graphic Toolbox. AlarmClient has some predefined functionality possibilities and its' parameters editing and configuration is accessible through Symbol Editor, what is a tool in ArchestrA IDE for graphical symbols modification as required.

Essential alarm manager component in ArchestrA IDE is graphical symbol named AlarmClient. The alarm mechanism in ArchestrA let to manage alarms triggered from itself, but also from alarm provider InTouch.

For adjustment of actual alarms signaling and manipulation it is necessary to enter alarms provider in AlarmClient. Alarm provider definition consists of alarm provider itself and group of generated alarms. General notation about setting up alarms provider is following:

- \provider!alarm_group
  or in more detailed form
- \provider!alarm_group!tagname

For example using generated actual alarms from InTouch the expression should be in this form:

- \intouch!$System

There are three different alarm types for configuration in ArchestrA IDE according to time of alarm generation [6]:

- current alarms – which presents actual real-time generated alarms,
- recent alarms – generated recently, usually a few minutes ago,
- historical alarms – triggered before a long time period.

In case of historical alarms there is a need to use a database for archivation of alarms generated longer time ago. For compatibility with Wonderware InTouch or Application Server is used Wonderware Historian in combination with Microsoft SQL Server to provide industrial database services.

Database of alarms in Historian is necessary to be created by tool Alarm DB Logger Manager delivered with InTouch installation. By means of this tool it is possible to set alarm provider (InTouch or ArchestrA), alarm groups and then we can create a database for alarms archivation (Fig. 2).

After configuration and starting a functionality of Alarm DB Logger Manager the alarm logging into the database is automatically provided.
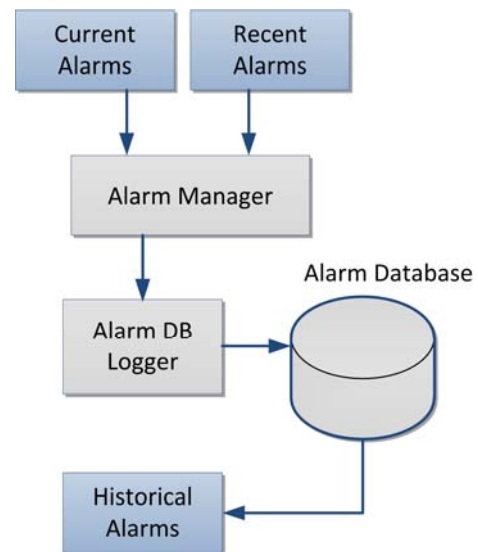
Fig. 2 Alarm system functionality scheme for generating Current, Recent and Historical Alarms. There is a need for coexistence with Alarm Database to gain historical alarming.

IV. APPLICATION EXAMPLE

A. Model Intelligent House

The Intelligent House model (Fig. 3) consists of two floors, which are rooms in the house [7]:

- In the room upstairs is realized temperature control with temperature sensor and heating device simulated by bulb. The room is equipped also with two fans to possibility of air circulating and thereby the change of temperature.

- In the room downstairs is realized control of lighting. There is placed lighting sensor, the source of light is bulb and light coming from surroundings through jalousie controlled by a stepper motor.

Fig. 3 Model Intelligent House

Essential controlled variables in the model Intelligent House there are lighting and temperature. These variables are technologically represented in controlled system by two bulbs in separated rooms. One of them perform lighting function and the second one is used like a heating device.

System control is provided by network LonWorks composed of sPoint modules and has next features:

- sensing lighting in the room,
- sensing temperature in the room,
- warm air exhausting,
- sensing lighting intensity,
- sensing heating intensity of heater.

Communication with higher level of control is implemented with Factory Suite Gateway.

B. Application

Visualization application of model Intelligent House (Fig. 4) was created using InTouch and ArchestrA software technologies and consists of main screen and secondary screens.

Fig. 4 Visualization application of controlled model Intelligent House created in ArchestrA IDE

In front of the screen in the part Historical is a button to invoke a window with historical alarms. On the main screen there is realized an alarm signalization light, which is blinking when is triggered unacknowledged alarm situation. Next to this light element is localized a button for showing the window (Fig. 5) with detailed list of acknowledged or unacknowledged alarms. Each alarm situation is displayed in one row and can be defined in different color rendering according to setting of alarm priorities. The area of alarm list is active, it means that there is a possibility to invoke a context menu by right-clicking on every line.

Fig. 5 Actual alarms window with the AlarmClient ArchestrA symbol

Context menu on each alarm line has several items and can be used to acknowledge of individual alarm or for adding a commentary and others. There is placed an external button for acknowledgment of all alarms under the area of alarm situations list.

V. CONCLUSION

The main benefit of this work is the practical implementation of ArchestrA alarm object in the supervisory control concept of laboratory model. Special ArchestrA alarm object is available only in new versions of Application Server and it has a different configuration and whole engine issues when comparing with InTouch alarm object functionality. ArchestrA alarm object utilizes alarm generation on automation objects supported by the functionality of specific application objects.

REFERENCES

[1] J. Flochová, T. Lojan, "Tool Designed for Supervisory Control of DES and Model-Based Diagnosis", In: *ITI 2009, Proceedings of the 31st International Conference on Information Technology Interfaces, Cavtat/Dubrovnik, Croatia*, SRCE University Computing Centre, University of Zagreb, 2009, pp. 33-34.

[2] L. Landryová, M. Babiuch, "Modelování objektů v průmyslových aplikacích (Modeling Objects of Industrial Applications)", Handbook of Research on Social Dimensions of Semantic Technologies and Web Services, Chapter XXXVI, Information Science Reference, 2009, pp. 743-759, ISBN 978-1-60566-651-8.

[3] D. Popescu, Q. Meng, A.A. Ionete, C. Ionete, R. Aguridan, "Remote vs. Simulated, Virtual or Real-Time Automation Laboratory", In *International Conference on Automation and Logistics, Shenyang, China*, August 2009, pp. 1410 – 1415, ISBN 978-1-4244-4794-7.

[4] J. Flochová, I. Zolotová, "Vizualizačné prostriedky, systémy SCADA/HMI", AT&P journal, Vol. 8, No. 12, 2001, pp. 28.

[5] L. Chrómčák, J. Koziorek, "Návrh aplikace s bezpečnostními prvky", Technická diagnostika 2010, Vol. 19, Ostrava, Asociace technických diagnostiků ČR o.s., 2010, pp. 30, ISSN: 1210-311X

[6] R. Mihaľ, "Alarmy, trendy a bezpečnosť v supervíznom riadení (Diplomová práca)", Košice: TUKE, FEI, 2010.

[7] T. Andrek, "Viacvrstvová automatizačná architektúra – priemyselný aplikačný server – laboratórny model Inteligentný domček (Diplomová práca)", Košice: TUKE, FEI, 2009.

Analyzing long tail visitors of web sites of different size

Adela TUŠANOVÁ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

adelatusanova@gmail.com

Abstract— this paper presents analyzes of so called long tail visitors of three different web pages. After brief introduction to long tail theory I present how to use Google Analytics to split visitors into keyword segments. In next section I demonstrate how to use these segments to analyze web traffic and I also compare long tail keywords of three different websites of different size. I used advanced segmentation to find out the best long tail keywords for all of three websites, to propose approaches for next optimization.

Keywords— long tail, Google Analytics, segmentation, web analytics

I. Introduction

Advantages and availability of internet is constantly increasing every year. There are thousands of new web sites with the same or similar topics, what means that companies started to fight for the visitors. The aim is to increase conversion rate, what is the ratio of visitors who convert into desired action on given web site. Conversion actions might include purchase, submitting a form, newsletter subscription, software download or other activities. Owners of web applications and developers realized the importance of keywords - they want to to have one of the best positions in Google search results on particular keywords or pay as little as possible for online advertising.

Keyword is a significant word or phrase, relevant to the web page or document in question [1]. User enters keywords to search engine and expects the most relevant pages which contain these keywords. Well optimized web site brings more visitors. It means that one of the most important processes in search engine optimization is proper use of keywords on the page.

One of the biggest problems for website owners or copywriters is to choose right keywords. There are two options: focus on major keywords or focus on bigger amount of particular keywords which are not so popular.

The main advantage of optimizing pages for the major keyword is usually the amount of searches per month. On the other hand, these keywords are usually more general and therefore the conversion rate, for example in case of e-commerce sites, can be much lower than would be in case of particular words.

Competition for major keywords is usually too high and therefore it is difficult to get the best positions in search engines. Also in case of PPC (pay per click) ads the price per click for major keywords is much higher than for particular keywords. These are the main reasons why become long tail keywords more and more popular.

II. Long Tail

The Long Tail was popularized by Chris Anderson in October 2004. The term is describing the retailing strategy of selling a large number of unique items in relatively small quantities – usually in addition to selling fewer popular items in large quantities [2]. In other words, products that are in low demand or have low sales volume can collectively make up a market share that rivals or exceeds the relatively few current bestsellers and blockbusters, if the store or distribution channel is large enough [3].

In web environment, long tail (see Fig 1) is represented by long tail keywords. Long tail keywords are keyword phrases with at least 2 or more words in them. They are usually highly specific.
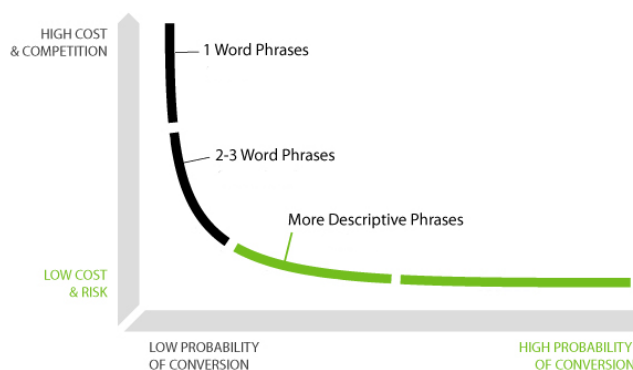
Fig 1. Long tail in web environment

Long tail keywords are cheaper and less risky. These keywords bring in customers looking for something specific. This will mean a much higher conversion rate because web site obviously has something that the customer is specifically looking for. Analysis of these keywords can help site owner to decide for which keywords (s)he should optimize and target his (her) website.

III. ANALYSIS OF LONG TAIL KEYWORDS IN GOOGLE ANALYTICS

There are many advanced statistical tools, which allow analyzing long tail keywords. One of the most popular is Google Analytics [5]. In order to segment and analyze long tail traffic it is important to set up advanced segments using regular expressions. These segments allow us to analyze historical data and also to trace the most popular phrases consisting of multiple words. Regular expressions are a set of key combinations that defines a pattern for a search to match. They allow us to create different segments for phases consisting of two, three or more words. Example of regular expression for segmentation of phrases consisting of two words is $^\wedge(\backslash S+)\backslash s(\backslash S+)\$$, where the characters have the following meanings [6]:

> $^\wedge$ beginning of chain
> \S any character except white characters
> + At least 1 time (max unlimited)
> \s "white" characters (space, tab, line break characters)
> $ The end of chain

This regular expression checks whether the string begins and ends by any character except "white" character. These characters are separated by a space, tab or line break character.

I have created four segments for our next analyzes (see Fig 2):

- Single-keyword queries
- Two-keyword queries
- Three-keyword queries
- Multi-keyword queries (more than three keywords)

Fig 2. Advanced Segments in Google Analytics

The advantage of Google Analytics is that once you create advanced segments you can access them across all of your website profiles, so there is no need to create it for each website profile separately. This was very handful for our analysis. I used these advanced segments to find out particular keywords for each website. The aim was also to compare traffics of three sites of different sizes and find out if the amount of long tail visitors depends on web site size or not.

IV. ANALYZING LONG TAIL VISITORS OF WEB SITES OF DIFFERENT SIZES

For our analyses I chose three different web sites – ubytovaniekosice.sk, qsystem.sk and h-profit.sk.

The first one, ubytovanie.sk, offers information about accommodation in Kosice. This website represents website with highest traffics in our case study. In last year (last 12

months) it had 16 326 visits with average 7,46 page views per visit. Almost 78% of visitors were new. Bounce rate (the percentage of visitors who enter the site and left rather than continue viewing other pages within the same site) was 37,91%. Almost 84% of visitors came via search engines (see Fig 3).
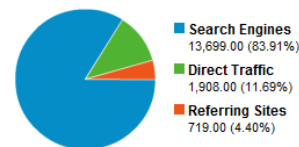
Fig 3 Traffic source overview - ubytovaniekosice.sk

The second one, qsystem.sk, offers information about company, which provides quality services and metrology. In last 12 months it had 4 107 visitors with average 2,55 page views per visit. More than 72% of visitors were new. Bounce rate was 51,33% what means that more than a half of visitors left website immediately. More than 76% of visitors came via search engines (see Fig 4).
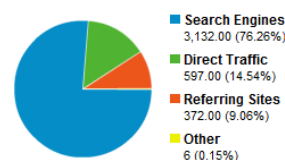
Fig 4 Traffic source overview - qsystem.sk

The third website, h-profit.sk, offers information about young and small accounting company. It had only 502 visits last 12 months with average 3,47% page views per visit. Almost 88% of visitors were new, but also almost 58% of visitors left website immediately. More than 63% of visitors came via search engines (see Fig 5).
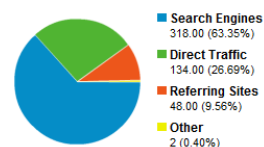
Fig 5 Traffic source overview - h-profit.sk

As we can see, most users who have visited analyzed website came just because the search engines. Therefore it's worth to analyze further keywords, especially long tail keywords which initiated these incomming visitors. I used advanced segments, as is described in chapter III. of this paper.

In case of website ubytovaniekosice.sk, only 5% of users used single keyword queries. Almost 62% of users used two keywords queries and 22% used three keyword queries. Rest 11% used multi keywords queries (see Fig 6).
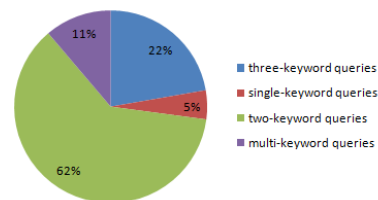
Fig 6 Segmentation of Long tail keywords – ubytovaniekosice.sk

In case of website qsystem.sk, 14% of users used single keyword queries, 50% of users used two keywords queries and 22% used three keyword queries. Rest 14% used multi keywords queries (see Fig 7).
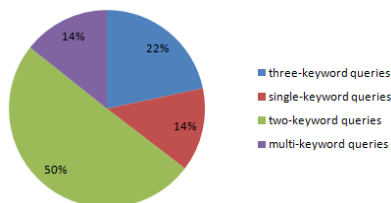
Fig 7 Segmentation of Long tail keywords – qsystem.sk

In case of website h-profit.sk, only 2% of users used single keyword queries, 79% of users used two keywords queries and 5% used three keyword queries. Rest 14% used multi keywords queries (see Fig 8).
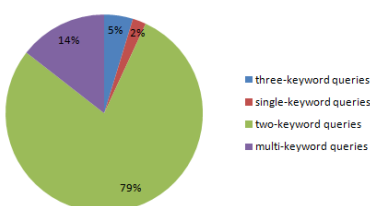
Fig 8 Segmentation of Long tail keywords – qsystem.sk

Comparing results according to website size, there are no significant differences - in all cases, two keyword queries are taking the biggest part of all queries. It is logical, because single keyword queries are too general. Therefore, if there is a lot of competitors for this single key word, user will maybe click on link on better positions as our result is. As we can see three and multi keyword queries are taking significant part of all queries also, therefore it is important to focus on these keywords during optimization of website or ads.

We can use our advanced segment to find out, which are these long tail keywords. In Google Analytics, we can find the best keywords for each segment. As an overview, we can use list of all long tail keywords, which include the best two, three or multi keywords (see table 1). Based on the data, website owners can optimize the site as well as advertising campaigns for long tail keywords and thus gain competitive advantage.

We can find some additional information in table 1. For example, users often search for different variations of company name. They also try to specify company location. As we can see, there are also a lot of phases which have the same meaning, but the words have different order or are with/without diacritics. This is specific for Slovak or Czech language and it is very important to keep it in mind during creating ad text and selecting proper keywords.

V. CONCLUSION

Based on an analysis of keywords of three websites of different sizes, we can say that independent on the size of the websites long tail keywords are important source of new visitors. These keywords represent the largest part of all the keywords that bring users to the website. Therefore, regardless of size of website, I highly recommend website owners to optimize the site as well as advertising campaigns for long tail keywords and thus gain competitive advantage.

ACKNOWLEDGMENT

REFERENCES

[1] Web analytics glossary [online] <http://www.netstrategies.com/web-analytics/web-analytics-glossary.html>
[2] Long Tail [online] <]http://en.wikipedia.org/wiki/Long_Tail>
[3] Long Tail [online] < http://opencontent.wgbh.org/report/glossary.html>
[4] Long Tail keywords [online] <http://www.articlemarketinghq.com/general/5-myths-about-long-tail-keywords/>
[5] Clifton, B.: Google Analytics. Brno: Computer Press, a.s., 2009. 334 s.: il. ISBN 978-80-251-2231-0
[6] Measure Your Long Tail Traffic With Google Analytics [online] < http://secretswede.net/seo/measure-longtail-traffic-google-analytics-mayday-update/>
[7] Tušanová, A.: **Chyba! Nenašiel sa žiaden zdroj odkazov.**. Bakalárska práca, KKUI FEI TU v Košiciach, 52s., 2008.

TABLE 1
LIST OF THE BEST 10 LONG TAIL KEYWORDS

	ubytovanie.sk	qsystem.sk	h-profit.sk
1	ubytovanie košice	q system	h profit
2	ubytovanie kosice	iso 9001:2008	h - profit, s.r.o.
3	košice ubytovanie	školenie metrológia	h profit kosice
4	chata ruzin	interný auditor	bezpečnosť pri práci
5	kosice ubytovanie	nástroje kvality	bozp v predajni protipožiarna ochrana
6	kosmalt kosice	skolenie metrologia	interná smernica evidencia dph
7	turistická ubytovňa k2 metropol	fmea priklad	ochrana osobných údajov mzdy
8	robotnícka ubytovňa košice	školenie fmea	podmienky zasielania intrastatu
9	ubytovna kosice	interny auditor	poradenske služby
10	turistická ubytovňa metropol	metrológia školenie	s.r.o zalozenie kosice

Anomaly Intrusion Detection System

Anton BALÁŽ

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

anton.balaz@tuke.sk

Abstract—**This paper deals with proposal of intrusion detection system to inform system administrator about potential intrusion in the system. Proposed architecture employs statistical method of data aggregation, that allows detection based on the knowledge of user activity aberrance in the computer system from learned profile representing standard user behaviour.**

Keywords—**intrusion, detection, anomaly, threats**

I. INTRODUCTION

The aim of a intrusion detection is to identify all intrusion attempts correctly and recognize activities that should not be tagged as intrusion. In this context, intrusion is understood to be integrity or resource accessibility violation, which means security policy violation. Systems detecting such activity is presented as intrusion detection system IDS. For its activity, IDS utilizes computer system and network features or data analyzed for a subsequent resolution on attack existence or process.

II. INTRUSION DETECTION SYSTEMS

Means of analysis is referred to as method used by intrusion detection system to determine an attack. There are two known analysis approaches [1]:

- detection per pattern,
- detection based on anomaly.

In the detection per pattern, system behavior patterns are identified, corresponding to known attacks. This method uses evaluation of frames and data in them, whereby it is possible to determine an attack. It is possible to detect system behavior patterns corresponding to known attacks from audit trails, logs (status records), or changes in attacked system. In the case of detection based on anomaly, each non-permissible anomaly from expected system behavior is identified. Expected system behavior is predetermined manually or automatically by prepared profile characterizing user/system behavior in the computer system [8]. Automatic elaboration of this profile is generated by a software collecting and processing data that characterizes system behavior in a time and forms statistically admissible behavior pattern subsequently.

III. DESIGN OF INTRUSION METHODS IN UNIX SYSTEM

The system controlling log records are based on record comparison with already known patterns presenting intrusion. It requires database upgrade of these patterns on each occasion of a new intrusion type detected, otherwise it is not discerned and revealed. Unlike such system, IDS is based on statistical data evaluation conception, capable of single attack detection, without obligation to upgrade it constantly [7].

However, statistical approach has its drawbacks. An attacker can learn behavior of eligible users and achieve the same behavior, what can lead to the states, when an attack would not be detected, eventually if an eligible user changes his behavior, false alarm will occur.

A. One-Layer IDS Architecture

The term of single-layer architecture means that all components, of which IDS consists, are part of it. Integrity and simple manipulation with the system present an advantage of the single-layer architecture.

In the way of architecture design, it was necessary to consider, which computer system IDS would be used on and who would be its administrator as well. In this case, single-layer architecture meant simpler administration and possibility of centralized regulation of system resources consumption.

Performance

An essential meaning and impact on the performance of intrusion detection system at the same time, have primarily these hardware components: processor, main memory, system bus, network interface card, and hard-disk. In general, the main question of the IDS design is property (value) selection, that IDS is supposed to consider and check as it works. For instance, they can include:

- disk mechanism access,
- processor utilization,
- computing memory utilization,
- hard-disk utilization,
- data integrity,
- battery utilization,
- average speed of data transfer in network,
- average size of network frames,
- duration of network connection,

The designed system is based on system parameter control, that represent utilization of processor, operating memory, and the number of running processes. They were chosen by reason of constant changes of their values, and because minimum activity changes in computer system are revealed by high changes of these values, and a quick and accurate identification of user behavior anomaly is made possible.

B. User Activity Profile

Through the profiles of users activity, subject (or group of subjects) behavior is described concerning certain object (or group of objects), what provides marking or description for regular behavior of these subjects and objects. Behavior is characterized through statistical values. During certain time

user_profile		
PK	id	INTEGER
	cpuval	NUMERIC
	actmem	NUMERIC
	proccnt	NUMERIC

Fig. 1: Database Model

interval, quantitatively measured values are represented by statistical value. Time interval can be allowed strictly e.g. minute, hour, day, week, or it can be a time between two control events e.g. between user login and logout to a system, program initialization and termination, or file opening and closing. Values of user activity profiles are obtained from control logs of system survey.

In the designed system, user activity profiles are represented as database tables and single profile parameters are represented as table columns. It is possible to create and delete profiles from the database, and switch among them only if system control is not running at the same time or if they are not used by the designed system in other way.

C. Data Gathering Method

For user behavior profile, it is necessary to learn behavior of proper user for being usable and ready for system controlling. Behavior learning represents data collection and gathering, presenting system parameters represented by processor utilizing, main memory utilizing, and a number of running processes. These data are stored in system folder /proc, concretely in virtual files:

- /proc/stat,
- /proc/meminfo
- /proc/loadavg

In fact, /proc directory is not a standard directory because by its exploration, it is possible to find that its files sizes is null. It represents all details about Linux system including information about the core, processes, and configuration parameters.

Data gathering is realized in period with duration adjustable by system user through the menu settings. The period contains function call for assemble data archiving. This function parameters are represented by functions, discovering current values of corresponding system activity characteristics. In this case, processor utilization is presented by its average usage in the time interval of one second and it is represented in percentage. Value of the operating memory is being loaded straight from a file, and it is set in kilobytes. Running processes represent all the processes existing in the system.

1) Data Logging: Data storing is being realized in one-second time period and it is provided by a function which parameters are functions determining current values of appropriate system activity characteristics. First evaluated functions are presenting parameters and their return values are stored in a database table. Data is constantly stored in a database.

Model of database table is displayed in Fig. 1. For column titles, values presenting processor utilization are preserved by the column cpuval, values presenting operating memory utilization are preserved by the column actmem, and values presenting the number of running processes are in column proccnt.

D. Data Processing

As systems are allowed to hold and treat extreme amount of data, it was necessary not to work with all the gathered data but with the data inevitable for the system operation. Account on that, determination of useless data, their discard was required. Allowing the fact that each learned value of system activity is unique and necessarily required for proper system functionality, two possibilities came into the account.

First possible alternative was data collecting in time intervals longer than one second. This possibility would provide storage of smaller data quantity, however, it would fetches back some risks. It could cause unintended disguise of intrusion into the system, eventually it could cause belated detection and warning on the intrusion.

An entire problem with amount of data was solved with the second alternative. It allowed data gathering in arbitrary long periodic time intervals. The base formed assumption that all values would be divided in groups and an average would be computed from values within the group. This was done to n-multiple reduce table capacity. User is allowed to change the number of values within the group in the settings of the IDS system. By default, an average is computed within triple groups of values. By setting, it is necessary to consider the less number of values within a group from which the average would be computed, the more exact would be detection of anomalous state and on the contrary, the greater would this number be, the less accurate would the detection be.

IV. Detection Mechanism of Designed IDS

Model of the designed IDS is based on functional assumption entailing that deviation of user activity in the system can be detected by a comparison of new observed values to strictly given ranges. Values indicating ranges are concluded by previous observation of user activities or other values of the same type. Fixed boundary consists from averages of values measured in the learning phase and deviation factor representing instability of these values between single measurements. New measurement is defined as anomaly if its value falls within interval of reliable values obtained in the learning phase of user behavior in the system, multiplied with the deviation factor. Model is also based upon hypothesis that system error utilization is accompanied by abnormal system utilization. Following these characteristics relating abnormal system utilization, violation of security rules can be detected.

V. System Activities Recording

User activity in a computer system in the designed system substituted by values of parameters described in chapter II.B. Concerning timeliness of the mentioned values, it is possible to divide them in two groups:

- values loaded from the database representative learned user behavior,
- values loaded during running of IDS system representative current user activity in operating system.

However, system testing through the IDS is based on data comparison representative learned sample user behavior with data presenting current user activity in operating system. In order to feasibility of the comparison, it is necessary to access data, type-equivalent with compared samples. The accessing is provided by cyclic working function determining and computing values of required system parameters that are

being stored in arrays representing samples for current activity of the system user. As it is necessary that sample filling cycle of the current activity is running during comparison as well, its running is provided by a separate thread. However, this can cause collisions, especially in case of comparison during filling the sample up with data. In this case, testing would be inaccurate, at worst, the intrusion wold not be detected. Solution of such collision was provided by using semaphores directing the whole process of the system monitoring and testing, and cloning data array representing loaded sample of the current system user activity. During filling the sample, comparison is in the phase of waiting on the contrary, tested sample cannot be updated if the comparison is in process. Mentioned thread and cloned array provide sample loading constantly by continual loading of values presenting the current system user activity. Tested sample is updated in bulk only if it is allowed by the semaphore.

VI. DATA COMPARISON

Main questions, needed to resolve with the design of intrusion detection system based on anomaly occurrence in the operation system, were:

- threshold level of intrusion detection,
- characteristics select for user activity monitoring.

Threshold level represents boundary value, overrun of that means possible intrusion, where it is necessary to record the event and warn the administrator about its occurrence. Value of the level can be different in various cases, so its fixed setting was ineffective and improper. For that reason, designed system offers possibility of modification of the value and its storing for the following utilization. As an enormous amount of values are treated in the system frequently, storing boundary values independently and constant accessing them, would be excessively computing-difficult. As a consequence, in architecture of the designed IDS, boundary values are computed following maximal allowed variance represented by its factor, that is allowed to be set by user. Single computation of maximal allowed value is realized directly during data comparison, so the necessity of other memory variables utilization, that would preserve the value is not needed.

Data comparison is based on comparing of two constantly updated samples, represented by value arrays. The first sample is represented by an array of values loaded from the database, presenting learned user behavior. The second sample is loaded during running of the IDS, concretely during the testing phase, and it contains values representing current user activity in operation system. Single operation of data comparison is realized by a function, in which directly at comparison possible exceed of maximal flowing variance from the average is computed and tested, representing learned user behavior.

VII. INTRUSION DETECTION

By reason of false alarm prevention to intrusion, data evaluation is realized exponentially, what is provided by counting of standard user behavior variance. That guarantees that intrusion at lower variances would not be indicated. It is ensured by the fact that lower variances would not exceed the learned value that represents standard user behavior. In case of user anomaly behavior that variances exceed 10% range, the value of variance summary is the reached level presenting standard user behavior. If user behavior remains anomaly or anomalous

in other way, the value of variance summary increases as the exceeded value determining boundary of average behavior value is multiplied by the factor of maximum flowing variance, attribute denoting intrusion is not annulled, and thereby the intrusion is indicated. Flowing variance is presented by the form that multiplies average of learned user behavior value.

After successful test of sample representing user activity, defined action is performed by the program following the predetermined attribute denoting an attack. In case of intrusion detected by an attribute, report is generated warning of the event administrator of the system.

Intrusion detection is recorded to system log of syslogd. System administrator is able to review system log. IDS system is enable to filter these logs and display them to a console.

VIII. FALSE INTRUSION DETECTION

If user use new software or by other meaning is his behavior changed, designed intrusion detection system detects new, unknown behavior which can be so different from learned one, that system will signal intrusion. It is necessary to remember user sequences, which is unexpected and can represent new user behavior which need to be added to patterns of expected behaviors. It is necessary to create new behavior profile or delete primary profile and create new profile to avoid inconsistency in profile names. This could represent hazard because if user's behavior is changed it is assumed that some activities will not be performed again.

It is necessary to count even an attack to system caused by user himself. User's rights are misused and his behavior matches entitled activity sequence. In that case intrusion will be not detected because user representing the attacker perform precise attack not violating border values of entitled behavior. Worse case represents situation when IDS system learned excess values of utilization of CPU and others system resources. In this incident attackers need to terminate running programs or decrease system resource utilization. After that it is possible to execute other software which utilization is boundaries of learned values.

It is possible to detect system activities which could be detected as false alarm. It is caused by e.g. operating system updates or applications updates with higher system utilization. Intrusion could by detected by executing unknown software with eligible user. Running programs not executed in phase of learning leads to fire detection system intrusion by unknown activities. The architecture of designed intrusion detection system is in the Fig. 2.

IX. EXPERIMENT

Concept of the experiment was to prove system functionality. The intrusion detection system included learned profile of standard computer system behavior. Time interval of learning profile was 1200 seconds. Size of the sample was 10. Error coefficient was 3 which enable system to detect anomaly from profile behavior. Number of values from which the average value was computed was 3. Activity messages was during learning phase and testing disabled. System testing of the designed system was realized in several steps representing different types of system attacks.
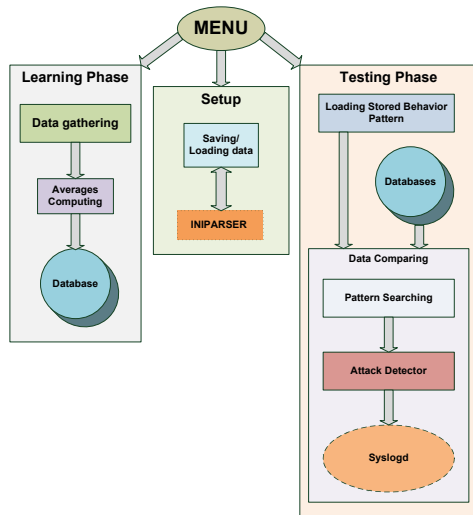
Fig. 2: Architecture of Designed IDS

TABLE I: IDS Results of Number of Requests

id	cpuval	actmem(kB)	proccnt
1	1,98	175249	48
2	16,23	175917	48
3	45,71	176237	48
4	74,82	176533	48
5	140,32	176825	50
6	270,02	177409	50
7	413,59	178521	50
8	622,10	179825	50
9	526,36	179073	50

A. IDS Detection to Anomaly System Behavior

For final testing of the purposed intrusion detection system was chosen experiment, which represents real IDS usage. Realization of this test requires existence of computer system as server which provides resources to its clients through Internet and computer system as client with goal server's denied of services. Client represents attacker.

B. Attack Realization

After successful learning phase, IDS system on the server begin to perform detection phase. After profile loading, server accept user requests. Attack on the server was performed by clients requests. During this attack it was recorded huge numbers of user requests. Even after 30 seconds of intensive testing, system does not detect intrusion.

C. Results

With normal user activities there was not detected intrusion or other malicious activities. During performing attack IDS system records increasing resources utilization. By exponential approach was ensured that attack was not early detected, but after some time, in this case after 12 seconds. Threshold limit represents value 75,83 and during attack identification system utilization value reached 140,32. Maximum system utilization with this attack gained value 622,10.

Measured values are in the Tab. I.

Graph of the measured values is in the Fig. 3.
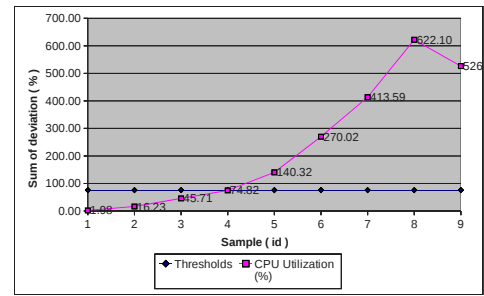
Fig. 3: CPU Utilization

X. CONCLUSIONS

With information and communication technologies progress, issue of operating system security has constantly higher importance. Information value of data increases, it is very important to prevent sensitive data compromising. This paper presented intrusion detection system and its implementation for the purpose of functionality verification of the designed architecture. Testing of user behavior anomaly using the method of statistical data evaluation. For better results, single testing phase was expressively fasten, what would entail excessive usage of system resources, having negative impact on system response from the user viewpoint.

ACKNOWLEDGMENT

REFERENCES

[1] Dorothy E. Denning: An intrusion-detection model,IEEE Trans. Softw. Eng., 1987
[2] Štefan Hudák: Reachability Analysis of Systems Based on Petri Nets, 1999, ISBN 80-88964-07-5
[3] Zhuowei Li and Amitabha Das and Jianying Zhou: Theoretical Basis for Intrusion Detection, 2005
[4] Vokorokos Liberios: Digital Computer Principles, 2004, p. 232, ISBN 9639548 09 X
[5] Vokorokos Liberios and Baláž Anton and Chovanec Martin: Intrusion Detection System Using Self Organizing Map, Acta Electrotechnica et Informatica, 2006
[6] Liberios Vokorokos and Alžbeta Kleinová and Ondrej Látka: Network Security on the Intrusion Detection System Level, INES 2006 10th International Conference on Intelligent Engineering Systems 2006
[7] Jingmin Zhou and Mark Heckman and Brennen Reynolds and Adam Carlson and Matt Bishop: Modeling network intrusion detection alerts for correlation,ACM Trans. Inf. Syst. Secur. 2007
[8] Rebecca Gurley Bace: Intrusion detection, Macmillan Publishing Co., Inc. 2000

An architecture overview of the smart-home system based on OSGi

[1]*Marek Novák* [2]*Miroslav Biňas*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]marek.novak@tuke.sk, [2]miroslav.binas@cnl.sk

Abstract—**Currently there are various ongoing European projects devoted to improvement of life of elderly people. They are aimed to provide higher degree of independence, autonomy and safety. The applications built for this purpose imply requirements on simple scalability of such systems and interoperability of hardware layer. In this paper we present our platform based on state-of-the-art software technique OSGi. We describe the architecture of the system mainly from a view of OSGi modules(bundles), OSGi4AMI middleware, data interchange standards as Hessian and JSON, OSGi web development frameworks and UI interfaces. We identify the advantages as well as limitations of used technologies. Finally we introduce future enhancements of the system.**

Keywords—**smart home, services, OSGi, ubiquitous computing.**

I. INTRODUCTION

According to demographics studies, which predict increasing number of people above 70 [1], European Union supports many research projects oriented on improvement of independent life and well-being of elderly and disabled people [2]. Technical University of Košice has participated in development, testing and deployment of a smart-home system for the *MonAMI* project [3]. The *MonAMI* system provides the functionality for local and remote monitoring and actuating devices located in the beneficiary's household and alert notification in case an unpredictable situation occurs.

In the process of MonAMI system design and implementation, which was done in cooperation with project's partners, we have identified several improvements and ideas how the system could be enhanced. Therefore, we have created our own branch of the system that we continue to work on, after the end of the project.

In this paper we would like to focus on the architecture solution rather than on functionality aspects. The sections in the paper are divided as follows: the first section is devoted to overview of the whole architecture, in the second section we describe decision making and rule definition system and third section describes user interfaces.

II. MONAMI SYSTEM ARCHITECTURE

The whole system is divided into three layers: hardware layer, core system and external services which may be located either on the same computer or may communicate remotely with MonAMI core. The core of the system is implemented in Java programming language, based on a modular framework *Apache Felix OSGi* [4]. Modules, in OSGi perspective called bundles, interact through interfaces defined by OSGi4AMI. Among external services there are: Email, SMS gateway, multimedia centre, etc.

A. OSGi4AMI middleware

Ambient Assisted Living (AAL) applications make use of the intermediate layer between device technology on one side and rule logic with user interfaces on the other side. The advantage is the flexibility and adaptability of future hardware technologies as well as new scenarios based on beneficiaries' requirements. Such approach helps also in a development process, when more different developer groups may work on a common application in parallel.

Several middleware platforms have already been and still are in a process of development in cooperative projects [5], [6], [7], [8]. Although there have been various middleware proposed, none of them have become a standard yet. They are developed for slightly different domains and designed under different goals, therefore their implementation vary. Though, the main idea is the same [8].

In our system we have used *OSGi4AMI* middleware [9]. The main idea is to hide hardware specific protocols and standards. For example, services do not have the information whether a temperature value came from a sensor built on a wired or a wireless technology. In Fig.1 there is shown a simplified scheme of *OSGiAMI* middleware. The hardware layer (in our case ZigBee and 1-Wire network) is propagated through abstract interfaces as TemperatureSensor, MotionSensor, etc. to the layer of services. The interfaces are implemented in Java programming language and serves also as connection points for bundles in OSGi platform. The upper layer, labelled as *External services*, represents services or applications that are not a part of OSGi platform.
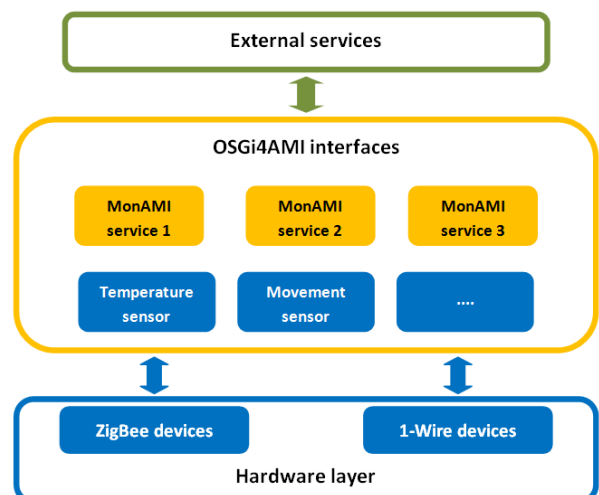
Fig. 1. Simplified scheme of OSGi4AMI middleware

B. OSGi platform

The OSGi technology [10] brings the modularity in Java programming language, that allows applications to be constructed from small, reusable and collaborative components. Each component, called bundle, has its own life cycle and may be remotely installed, started, stopped, updated and uninstalled without requiring a reboot of the system. The installation of new modules and their management is therefore very easy. The bundles must define only dependencies between them through interfaces, the searching and binding of required services is handled by OSGi framework. The current OSGi specification version 4.2 is developed by the OSGi Alliance. There are available several OSGi implementations as Knoplerfish, Apache Felix, Equinox and Concierge OSGi [8].

Although the OSGi technology provides many advantages as security, modularity, scalability and is used in several AAL projects [11], [8], [12], we have identified following limitations:

- **Thread safety** - Coming from a nature of used sensor technologies, each device was running in an own thread. OSGi technology do not solve concurrent access to shared services. For example, two different gas sensors may ask an alert service to produce an alarm on the same time. The developer must be aware of this possibility and have to ensure execution of the service in transaction.
- **Import-Package complexity** - With the increasing number of used packages from other bundles raises the complexity needed to define imported packages in manifest file.
- **Web frameworks support** - The support for some technologies as Java Server Pages (JSP) or Java Server Faces (JSF) is not direct and there are needed third party bundles, providing this functionality.

Spring Dynamic Modules [13] seems to be a suitable solution solving first two mentioned drawbacks of the three. In the future we are going to integrate this technology in our system.

C. System components

Global architecture overview of the system is depicted in Fig.2. Residential gateway (RG) represents arbitrary computer with the support of Java Virtual Machine (JVM) and sufficient number of USB ports to connect device networks. There are located all OSGi bundles: driver for each physical device technology, configuration bundle, logging support, HTTP server Jetty, Hessian binary protocol server for HTTP communication with external applications and rule management bundles.

Beneficiary may interact with the system in different ways: by smart-phone with Android operation system, through multimedia centre via TV situated in user household or remotely by web user interface. Carer is notified about unexpected situation by SMS and Email alerts or directly by a phone call.

III. RULES DEFINITION

Actions realized by the system are activities triggered when events from sensors fulfil defined rules (conditions). Behaviour of majority of services can be simply defined by following formula: *if values from sensors exceed thresholds take an appropriate action*. An example is a *Wake up service* which sends an SMS to carer in case no motion is detected in a predefined time interval.
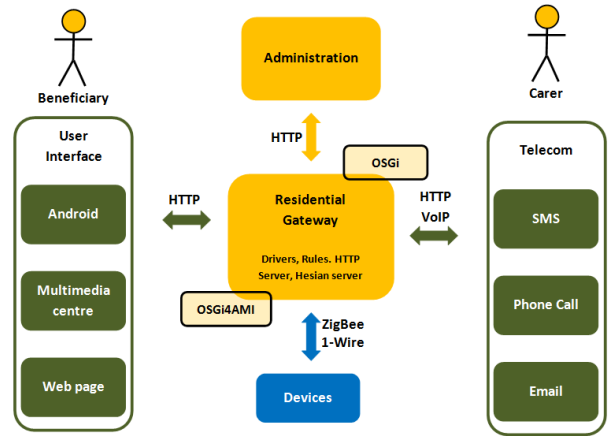
Fig. 2. Main components of MonAMI system

We have used two different approaches for rules definition:

- Conditions implemented directly in source code of the bundle.
- Domain specific language for service configuration based on XML.

Domain specific language (DSL) significantly simplifies the process of new service definition, where rules are defined in one configuration file. In the first approach, however, it is needed to implement new bundle with specific source code for each rule separately. We have used XML format because of its advantages as: hierarchical structure of final configuration file, existence of validators and existence of APIs parsing the XML file [14].

Below is listing of an example rule definition for Wake up service. Tag <*sensor*> defines a sensor from which events are evaluated in the particular service. Consecutive action comprises of two parts: condition (defined by <*condition*> tag) and taken action (defined by <*action*> tag). The events are generated by sensors or a timer.

```
<service>
    <name>WakeUPService</name>
    <type>PersonPresenceDetector</type>
<sensor>
    <name>motion</name>
    <type>MotionSensor</type>
    <location>Bedroom</location>
</sensor>
<condition>
    <and>
    <cond oper="FALSE" val="TRUE" duration
        ="3600">motion</cond>
    <cond oper="ht" val="08:00:00" >time</
        cond>
    </and>
</condition>
<action>
    <actIf val="ON">sms</actIf>
</action>
<service>
```

Our approach is similar to that of T. Fuxreiter et al. [8]. They have used finite state machines (FSM), where the actions are: setting a system variable, starting a timer, controlling an actuator and the events are incoming messages from sensors

or by timers. Both, FSM and DSL provides abstraction from service information for service developer. We see the strength of XML service definition in further enhancement of the system, where the user will be able to define services only by some "clicks" in graphical UI.

IV. User Interfaces

As a main user interface has been selected *Universal Control Hub* (UCH) middleware with a web application built on top of it [15]. UCH seems to us very complex solution requiring high input effort with small added value in a sense of scalability and future enhancements. Therefore, at Technical University we have created our own UI, called *TUKEamivue* built on JSP and Ajax technologies and running on HTTP Jetty container. When implementing TUKEamivue service we have strived with several limitations of the OSGi framework.

A. Universal Control Hub

UCH middleware was developed within the scope of the EU FP6 project *i2home* [16] and is based on the *Universal Remote Console (URC)* Standard ISO/IEC 24752. Core part of the UCH is socket layer hosting discovery and control protocol between RG and UCH. The Socket layer is based on HTTP and provides following main messages: discovery, notify service, send request, send message.

The communication between UCH and RG is schematically depicted in Fig. 3. Generic communication service (GCS) is an OSGi bundle responsible for handling HTTP requests and responses on the side of RG. The events are passed from UCH to web based UI through HTTP. There are used different interfaces for beneficiary and carer. A temperature value measured by a sensor connected to the RG is passed through GCS to UCH and then to a web page built on a Webclient JavaScript Library (WJSL). We identified the UCH middleware redundant, since OSGi4AMI middleware hides device specific technology. Another deprival of the UCH is the posibility to send requests only one way. The events are created on the side of RG, however only the UCH is able to send request to RG. Therefore the UCH has to ask RG for a new values in small time intervals, which increase the communication between the two components.
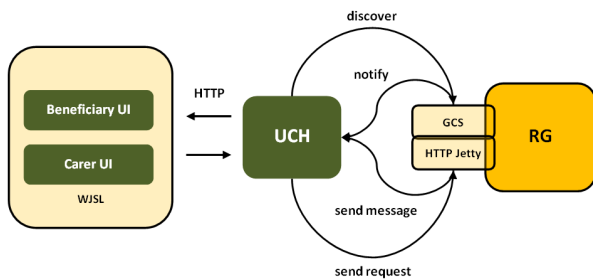
Fig. 3.   UCH communication middleware

B. TUKEamivue service

TUKEamivue service provides web UI for monitoring (Fig. 4), controlling devices located in a household, calendar tasks management and some additional functionality as weather forecast or interface for enabling and disabling system services. TUKEamivue is OSGi bundle running on RG providing

remote access only to one household. The service is based on JSP technology. Here comes the drawback of OSGi platform, since Apacha Felix do not provide bundles for JSP support. Also web application deployment is quite complicated, since OSGi requires servlet and resource file registration.

We have used Pax Web bundles provided by Open Participation Software for Java (OPS4J) community [17]. Pax Web extends OSGi Http Service with better servlet support, filters, listeners, error pages and JSPs and adds the support for easy installation of WAR bundles. Pax Web is still in development, so the users of this framework have to be aware of possible problems connected with bundle deployment.

Fig. 4.   Web user interface built on JSP and Pax Web framework

C. XBMC and Android user interface

Our aim is to provide more control options for the users, to allow them monitor the household through TV and media centre or smart-phones. We have developed the plug-in into XBMC media player centre [18] and Java application for controlling and monitoring a smart-home remotely through a smart-phone.
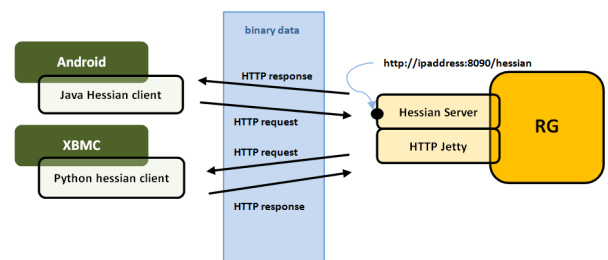
Fig. 5.   Communication between RG and UIs based on Hessian binary protocol

The main aspect was to support the communication between RG and XBMC or Android. We wanted to provide a common communication protocol for both platforms. Among the the possibilities were: web services, lightweight text-based open standard JavaScript Object Notation (JSON), building our own protocol or binary web service protocol Hessian. As the best solution for our requirements we have evaluated *Hessian*, because it provides mechanism for direct methods call in source code in both sides of client and server. Hessian is a binary protocol [19] well-suited to sending binary data without any need to extend the protocol with attachments. There are supported various programming languages for Hessian through third party implementations (C++, .NET C#, Perl, PHP, Ruby, Objective-C, etc.) Other possibilities as JSON, web services

and our own protocol we have drop out because they required additional development. From the behaviour of our application is it also obvious that frequent marshalling and unmarshalling simple data in interchange format as Web Services Description Language (WSDL) will be vain use of resources.

Communication in Hessian is defined by an interface on both sides - client and server. Server binds the object instance on the server side and client remotely calls exposed methods (Fig. 5). The consistency is assured by the "agreement". Hessian do not use any formal mechanism to check the consistency between the client and the server interface definition. If the consistency would be broken (e.g. server interface defines method with two parameters, and client is calling method only with one parameter)the on the server side will be thrown.

Fig. 6.   Final solution with Asus touch screen, sensors and care phone unit

V. Conclusion

The OSGi technology provides highly suitable framework for smart-home application development. Among the main advantages are modularity, scalability and dynamic deployment and management of system modules. Since the technology is rather new and in fast development, the support for external frameworks is sometimes small. However, the situation is steadily improving thanks to communities like *OPJ4S* and *Spring Source*. OSGi current development is focused on the connection of distributed OSGi frameworks as well as connection to non-OSGi remote services. The recent OSGi specification 4.2 addresses these issues and first implementations are already under development [20].

We are currently focused on development of a configuration web interfaces for OSGi services based on Vaadin framework [21], service definition UI with aspects of user behaviour detection. Current research is oriented on activity classification, mainly user tracking in household environment (movement and device usage) and finding anomalies in user behaviour. Our aim is to create a dynamical model that will adapt to changing conditions in the user behaviour by the knowledge of some dependencies between the activities. For this purpose we plan to integrate techniques like Dynamic Bayesian Networks (DBN) or Hidden Hierarchical Markov Model (HHMM) into our system (Fig. 6).

Acknowledgment

References

[1] A. Zaidi, "Features and Challenges of Population Ageing: The European Perspective," March 2008. [Online]. Available: http://www.euro.centre.org/data/1204800003_27721.pdf

[2] AAL, "Overview of Funded and Running Projects Call 1 (2009) and Call 2 (2010)." [Online]. Available: http://www.aal-europe.eu/projects/aal-brochure-2010

[3] D. Šimšík, J. Bujňák, D. Siman, and M. Novák, "Experience with ambient services for seniors in slovakia," *Lékař a technika: biomedicínské inženírství a informatika*, vol. 40, no. 2, pp. 36–40, 2010.

[4] A. Felix, "Apache Felix - Overview," accessed: March 2011. [Online]. Available: http://felix.apache.org/

[5] M.-R. Tazari, F. Furfari, J.-P. L. Ramos, and E. Ferro, "The PERSONA service platform for AAL spaces," 2009.

[6] A. Pitsillides, E. Themistokleous, G. Samaras, and O. M. Winnem, "Overview of mpower: Middleware platform for the cognitively impaired and elderly," in *IST-Africa 2007 Conference*, May 2007.

[7] A. Hein, M. Eichelberg, O. Nee, A. Schulz, A. Helmer, and M. Lipprandt, "A service oriented platform for health services and ambient assisted living," in *Advanced Information Networking and Applications Workshops, 2009. WAINA '09. International Conference on*, May 2009, pp. 531 –537.

[8] T. Fuxreiter, C. Mayer, S. Hanke, M. Gira, M. Sili, and J. Kropf, "A modular platform for event recognition in smart homes," in *e-Health Networking Applications and Services (Healthcom), 2010 12th IEEE International Conference on*, 2010, pp. 1 –6.

[9] A. Marco, R. Casas, G. Bauer, R. Blasco, A. Asensio, B. Jean-Bart, and M. Ibañe, "Common osgi interface for ambient assisted living scenarios," in *Ambient Intelligence and Smart Environments: Behaviour Monitoring and Interpretation – BMI - Smart Environments*, 2009.

[10] O. Alliance, "About the OSGi Service Platform," 2007. [Online]. Available: http://www.osgi.org/wiki/uploads/Links/OSGiTechnicalWhitePaper.pdf

[11] H. S. Kim and J. H. Son, "User-pattern analysis framework to predict future service in intelligent home network," in *Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on*, vol. 2, 2009, pp. 818–822.

[12] Z. Lei, S. Yue, C. Yu, and S. Yuanchun, "Shsim: An osgi-based smart home simulator," in *Ubi-media Computing (U-Media), 2010 3rd IEEE International Conference on*, 2010, pp. 87–90.

[13] SpringSource, "Spring Dynamic Modules for OSGi Service Platforms," accessed: March 2011. [Online]. Available: http://www.springsource.org/osgi

[14] M. Novák, "Easy implementation of domain specific language using xml," in *SCYR 2010 : 10th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice*, vol. 2, 2010, pp. 298–301.

[15] URC Consortium, "Universal Control Hub Reference Implementations," accessed: March 2011. [Online]. Available: http://myurc.org/tools/UCH/

[16] i2home, "Intuitive interaction for everyone with home appliances based on industry standards," accessed: March 2011. [Online]. Available: http://www.i2home.org/

[17] OPS4J - Open Participation Software for Java, "Pax Web," accessed: March 2011. [Online]. Available: http://wiki.ops4j.org/display/paxweb/Pax+Web

[18] XBMC, "About," accessed: March 2011. [Online]. Available: http://xbmc.org/about/

[19] C. T. Inc., "Hessian 2.0 Serialization Protocol," accessed: March 2011. [Online]. Available: http://hessian.caucho.com/doc/hessian-serialization.html

[20] R-OSGi, "R-OSGi - Overview," accessed: March 2011. [Online]. Available: http://r-osgi.sourceforge.net/

[21] Vaadin Ltd, "Vaadin," accessed: March 2011. [Online]. Available: http://vaadin.com/home

An Industrial Information Portal based on Virtualized Web Server

Miloš PAVLÍK, Peter KARCH, Roman MIHAĽ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

milos.pavlik@tuke.sk, peter.karch@tuke.sk, roman.mihal@tuke.sk

Abstract— **Virtualization is becoming a standard information technology practice. Virtualization solutions are being broadly used in development and production environments, providing tremendous benefits like scalability, flexibility and availability. The need for high availability for industrial information portal is clear and based on this fact, we used a virtualization technology to virtualize a web server, on which runs a production information portal. To achieve high availability of industrial information portal, we used high availability strategy and Hyper-V Guest Clustering scenario of Failover Clustering under Microsoft Windows Server 2008 R2.**

Keywords — **Virtualization, Web Server, Industrial Information Portal, Wonderware,**

I. INTRODUCTION

Virtualization brings many benefits not only in to the IT (Information Technology) environment but also in cybernetics and SCADA/HMI (Supervisory Control and Data Acquisition/Human machine interface) systems. Taking into account these benefits, we decided that we will use this technology in the effort of high availability of industry information portal. To achieve this goal we used the High Availability strategy and we have chosen Hyper-V High Availability implementation, specifically Hyper-V Guest Clustering. Guest Clustering (a cluster in a virtualized environment) provides two key values: Application Health Monitoring, which enables health monitoring of applications running within a VM (Virtual Machine) and Application mobility, which allows applications to failover from one VM to another VM.

II. WODERWARE INFORMATION SERVER - INDUSTRIAL INFORMATION PORTAL

Wonderware Information Server (before known as Suite Voyager) is an information portal for concentration and analysis of information. It uses the latest software technologies such as XML, VML, SOAP, .NET and *ArchestrA* technology platform. This portal is able to obtain all kinds of data, which may occur in an industrial enterprise. It allows to sort, organize these data, make them compact and provide them to the thin clients. As thin client can be used web browser Microsoft Internet Explorer, because of ActiveX components. This application thus becomes a universal client application for comfortable access to live data, alarm notifications, historical data and other information from the database applications and from the ongoing technological processes. Clients may have a graphical view of the current status of the process including the dynamic animation. With appropriate permissions can remotely interfere with the ongoing technological processes [1], [2], [14].

III. VIRTUALIZATION AND HIGH AVAILABILITY

A. Virtualization

Virtualization is the creation of a virtual version of something, like an operating system, a server, a storage device or network resources. Divided the hard drive into different partitions is some kind of virtualization, so it isn't so unknown term. So virtualization in its core is involved with adding "layers of abstraction" to the IT environment. These layers decouple one layer from another.

Virtualization has many benefits. As we said, if the applications use the same server for behavior, they can invade each other, for example because of updating or one of the applications CPU (Central processing Unit) usage is at hundred percent, etc. The risk is greatly reduced by dividing the applications onto different virtual computers. Virtualization gives flexibility, so the operating system does not depend on particular piece of hardware. There are no changes needed for, for example moving the virtual computer. It ensures high scalability of systems. It is not necessary to build a 'final solution' because systems can be expanded on the fly, without prejudice to operate services. By using virtualization technology, it's very easy to backup the systems and it's great to simulate technically difficult solutions without need to own or ensure additional hardware [3].

B. High Availability

High availability solutions provide ways to manage planned and unplanned outages. Some examples of planned outage include the installation of the operating system or updating application that requires server offline. Unscheduled downtime can only be the result of hardware component failure, or loss of physical servers due to natural disasters.

Server virtualization provides a solution for organizations

to implement a dynamic, flexible basic infrastructure, which minimizes the number of deployed physical servers, increases the use of resources and reduces long term operating costs. Removal of several physical servers to a common virtualization host requires a wider high-availability strategy as in traditional infrastructure. On the hardware level, the deployment of virtualization hosts on platforms that contain redundant or hot-swap components (resources, processors and memory) reduces the risk of unplanned shutdowns. With Windows Server 2008, is also the possibility to use the integrated features for managing **Failover Clustering** unplanned and planned outages. In Windows Server 2008 with Hyper-V as virtualization platform, we can integrate Failover Clustering as part of its high-availability strategy for virtualized infrastructure. Windows Server 2008 Failover Cluster is composed of at least two servers (nodes) that are interconnected through various network links, one to track the status of individual nodes. Every Failover Cluster node is connected to the common disk array, such as Storage Area Network (SAN), and only one node in the cluster can have custom settings for network and disk resources associated with the application or service at one time. There are two different levels at which we can implement failover cluster on Hyper-V virtualization at the host and the guest operating system level. There are two Hyper-V High Availability implementations, Hyper-V Guest Clustering and Hyper-V Host Clustering [4].

Guest Clustering provides two key values:

Application Health Monitoring – Enables health monitoring of applications running within a VM. For example; if SQL Server is placed in a VM then the health state of SQL will be monitored to ensure it is functional and take recovery actions if the SQL instance were to crash or hang.

Application mobility – Allows applications to failover from one VM to another VM. For example; if you need to patch the guest operating system you can move the application to another VM and reduce downtime.

It is supported by Microsoft to run Failover Clustering in a virtualized environment with Windows Server 2008 R2 Hyper-V; however the support policy varies for different guest OS versions [15].

IV. GOALS AND ANALYSIS

An industrial information portal – Wonderware Information Server from company Wonderware corp. represents dynamic portal type sites that acquire, combine and process the business data from different providers (not only SCADA/HMI applications and historical data but also data from MES (Manufacturing Execution System), ERP (Enterprise Resource Planning) applications). This data are organized and customized for needs of users and they are accessible to clients through universal portal interface.
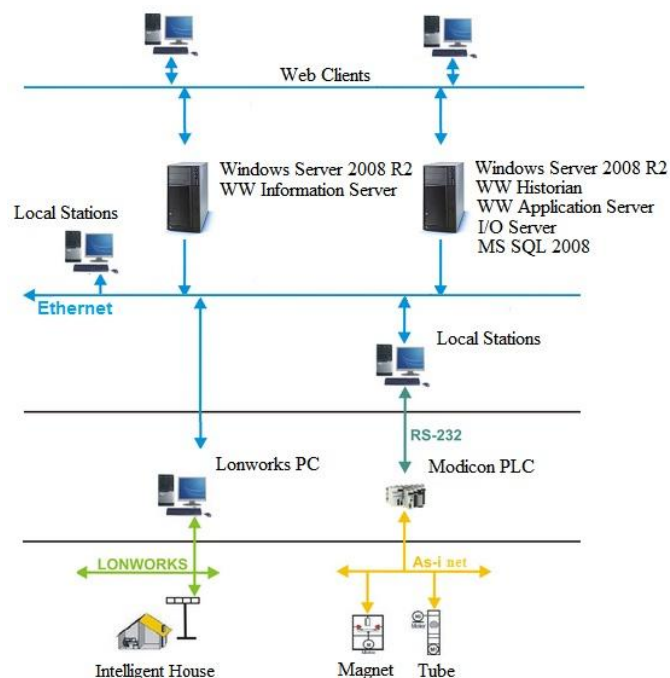
Fig. 1 Architecture of Distributed Control System in Laboratory

Wonderware Information Server is a tool that provides opportunities for remote visualization and process control in real time. There are few real controlled models in laboratory (intelligent house, tube, magnet) and also a few virtual-simulated models. All of these models are for research, presentation and educational purpose. Through information portal we have direct access for control of these models. Basically it allows access of visualized applications of these models, which were created in visualization software environment Wonderware InTouch, to an industrial information portal. At figure 1 is architecture of distributed control system at laboratory L509.

It is necessary that industrial information portal is always available. So basically there is need to apply a High Availability strategy on our system. There are few providers that provide the ability to apply this strategy. We have chosen Hyper-V High Availability implementation, specifically Hyper-V Guest Clustering. In a Guest Cluster implementation, the clustering service runs on the guest operating system in the virtual machine. This type of clustering provides high availability for applications that are hosted within the virtual machines.

Clustering within guest operating systems doesn't depend on clustering being enabled on the Hyper-V hosts. The requirements for clustering in guest VMs are the same as for physical machines operating in a cluster - suitable network connectivity and, depending on the services, access to shared storage. In the case of guest VMs, shared storage means iSCSI storage.

If there is need to move VMs between Hyper-V hosts with no downtime, we need to use Quick Migration or Live Migration technologies, which require clustered Hyper-V hosts.
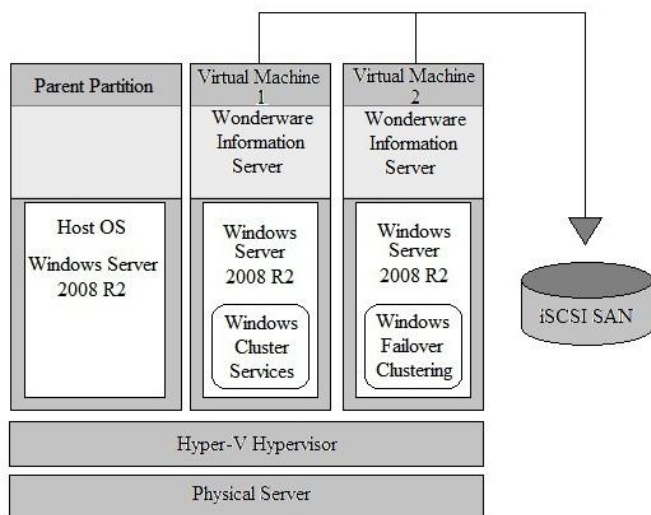
Fig. 2 Hyper-V Guest Clustering Within a Single Physical Server

In Figure 2, virtual machines 1 and 2 are running on the same server. In this scenario, the workloads being managed by the clustering service in the guest operating systems in the VMs will not survive a physical server outage or a crash of the parent partition. There are other scenarios of Failover Clustering, that are truly and more High Available but this scenario is very interesting for testing, training or demonstrations. In this case, shared storage must be an iSCSI SAN (Internet Small Computer System Interface Storage Area Network).

This scenario and High Available strategy is making the Wonderware Information Server (web server) more available and more resistant to unplanned outages and allows managing planned outages of web server.

V. CONCLUSION

Using the Failover Clustering and High Availability strategy for industrial information portal, was a reasonable demonstration to ensure a high availability of web server for the smooth running of an information portal. In this case we dealt with the scenario of failover clustering, where was used a single physical server, in which were two virtual machines with Wonderware Information Server, using shared data store. There is a strategy with higher High Availability, but for our purposes (testing, training and demonstrations) suffice this one.

ACKNOWLEDGMENT

REFERENCES

1. P. Karch, "Local and Remote Supervisory Control of Systems – Intelligent House, Traverse", Diploma thesis, Košice: FEI TU in Košice, 2008.
2. L. Takáč, J. Liguš, J. Sarnovský, " Systematic Design of the Web Portals" in *Proceedings of 19th EAEEIE Annual Conference on Innovation in Education for Electrical and Information Engineering*, Tallinn, Estonia 30. June – 2. July, 2008.
3. M. Pavlik, I. Zolotová, R. Hošák, L. Landryová, " Benefits of Virtualization in SCADA/HMI Systems" in *Procedings of the 14th International Conference on System Theory and Controll,* October 17-19, 2010, Sinaia, Romania, ISSN 2068 – 0465, pp. 379-385.
4. D. Clitherow, M. Brookbanks, N. Clayton, G. Spear, "Combining High Availability and Disaster Recovery Solutions for Critical IT Environments," IBM Systems Journal 47, No. 4, 563-575 (2008, this issue).
5. A.G. Vicente, I.B. Munoz, J.L.L. Galilea, "Remote Automation Laboratory using a Cluster of Virtual Machines", Del Toro, P.A.R. 2010 IEEE Transactions on Industrial Electronics 57 (10), art. no. 5409652, pp. 3276-3283.
6. M. Babiuch," The Usage of the New Technologies at the Education at the Department of Control Systems and Instrumentation", 2006, pp. 7-12. ISSN 1210-0471. ISBN 80-248-1211-8.
7. Božek, Pavol - Barborák, Oto: Key benefits of virtual technology. in *Intellektuaľnyje sistemy v proizvodstve*. - ISSN 1813-7911. - Vol. 15, No. 1 (2010), s. 115-119.
8. Jason S.K. Lau, George Q. Huang, K.L. Mak, "Web-based Simulation Portal for Investigating Impacts of Sharing Production Information on Supply Chain Dynamics from the Perspective of Inventory Allocation", Integrated Manufacturing Systems, Vol. 13 Iss: 5, 2002, pp.345 – 358.
9. J. Liguš, I. Zolotová, P. Karch, J. Ligušová, "Information and Control System of Traverse and its Integration into Cybernetic Centre", Electronics and electrical engineering. 2010, No. 6, p. 147-152, ISSN 1392-1215.
10. L. Landryová, P. Osadník, M. Valas, O. Winkler, "Integrating Tools and Methodologies in FutureSME Delivery Platform", Seminar ASR '2010 "Instruments and Control" © 2010, VŠB-TUO, Ostrava, ISBN 978-80-248-2191-7, pp. 199-205.
11. B. Sobota, J. Spišák, M. Straka, M. Gofčík, "Model výrobného procesu a technológie virtuálnej reality", Automatizace. Ročník 43, č. 3, 2000, Česká republika, ISSN 0005-125X, s. 186-188.
12. D. Popescu, Q. Meng, A.A. Ionete, C. Ionete, R. Aguridan "Remote vs. Simulated, Virtual or Real-Time Automation Laboratory" in International Conference on Automation and Logistics, Shenyang, China, August 2009, ISBN 978-1-4244-4794-7, s. 1410 – 1415.
13. J. Koziorek, "Education of Industrial Control Systems at VSB-Technical University of Ostrava", Czech Republic. in *CONTI'2008 - The 8th International Conference on Technical Informatics*, 5-6 June 2008, Timisoara, Romania. Vol.2, p. 135-140. ISSN 1844-539X.
14. http://global.wonderware.com/EN/Pages/WonderwareInformationServer.aspx
15. http://www.devx.com/MS_TechNet/Article/41725/9290
16. http://technet.microsoft.com/en-us/library/cc732181(WS.10).aspx

Authentication of mobile nodes by using EAP protocol

[1]*Milan VRABEĽ,* [2]*Peter JAKUBČO,* [3]*Peter FANFARA*

[1]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]milan.vrabel@tuke.sk, [2]peter.jakubco@tuke.sk, [3]peter.fanfara@tuke.sk,

Abstract—*In paper presented proposal is discussing about mobile communication systems for transferring high amount of data with high rate transmission such as UMB system are. It describes methods for authentication and authorization of mobile node in mobile communication network. Proposed method offers effective device authentication and user authentication during an initial call setup and as well as for mobile communication network in mobile communication system. Process for authentication and security ensure is performed by appropriate implementation of Extensible Authentication Protocol (EAP) in mobile communication network without using PPP. The next aspect of presented method is showing, that the performance of device authentication and user authentication is more securely and more efficiently, even when network nodes responsible for controlling of signaling for a mobile nodes, are logically or physically separated.*

Keywords—**UMB, EAP, Authentication**

I. INTRODUCTION

Since the mobile communication system typically operates via a Point-to-Point Protocol (PPP), a Challenge Handshake Authentication Protocol (CHAP) or a Password Authentication Protocol (PAP), which is framework that can work above PPP, is used for user authentication or device authentication. However, these protocols are not viable for an Ultra Mobile Broadband (UMB) system developed by the 3GPP2 to transmit more data at higher rates. Hence, an authentication and security technique that can support the UMB more efficiently is needed [1].

In mobile communication system such as 3rd Generation Partnership Project 2 (3GPP2), Code Multiple Access 1x (CDMA 1x) and Evolution-Data Only (EV-DO), a Base Station (BS) is responsible for managing radio resources and network entity within a core network [3].

Conventional authentication and security technologies for a 1xEV-DO system are not effective in a perfect protection against channel hijacking and allows for unauthorized use of a service without payment of a lawful charge for the service. Moreover, the conventional system is vulnerable to denial of a service caused by a message attack at a protocol level as well as at a Radio Frequency (RF) level. Accordingly, there is a need for system and a communication network that enable secure communications.

II. PROPOSAL PREDISPOSITIONS

A. CDMA systems

Very briefly remind, what in point of fact the CDMA is and differences with GSM. The CDMA is for communication canals division using cryptography by pseudorandom code. It exists in different modes for 2nd and 3rd networks generations. The GSM is using canal time-division method (TDMA).

The oldest standard – CDMAOne belongs to 2nd generation networks and was originally dedicated especially for voice communication. Different revisions increased step by step the speed of data flow to 64 kbit/s. Today most of telecommunication operators in America and Asia are migrating from CDMAOne technology to CDMA2000. The CDMA2000 technology includes different standards covered by 3GPP2, which is CDMA organization part for mobile networks development 3rd generation (3GPP- The 3rd Generation Partnership Project).

Today covering these modes:

- *CDMA2000 1xRTT* (voice + data to 144 kbit/s or to 307 kbit/s),
- *CDMA2000 1xEV-DO* (just data, DO=Data Only, to 2,4 Mbit/s or to 3,1 Mbit/s in revision A + voice over internet support) and
- *CDMA2000 1x EV-DV* (voice + data to 3,1 Mbit/s) [2].

B. EAP protocol definition

The EAP protocol is an authentication framework frequently used in wireless networks and Point-to-Point connections, defined in IEEE 802.1x. Protocol is encapsulating frames on 2nd layer. It can be used for authentication, authorisation and accounting (AAA) and uses services from RADIUS sever for key producing (per-session, per-user). The Base Station is as the first associating with Access Point (AP) (using open system authentication) and then EAP authentication starts. When EAP authentication finished successfully, the AP and BS have number of required keys for secure communication between them. Problems can have possibly only first phase, which is not secured and is vulnerable for DoS attacks [4].

There are several protocols for authentication:

LEAP

The Lightweight Extensible Authentication Protocol (LEAP) is a proprietary EAP method developed by Cisco Systems. LEAP uses a modified version of MS-CHAP, an authentication protocol in which user credentials are not strongly protected and are thus easily compromised. The user mane here is not encrypted and it uses just 16 bits DES algorithm. This method was not build for networks with free access to data medium.

EAP-TLS

The security of the EAP-Transport Layer Security protocol is strong, provided the user understands potential warnings about false credentials. It uses PKI to secure communication to a RADIUS authentication server or another type of authentication server. Both neighbours are authenticating each other by using a certificate: first the authentication server send crated certificate to a client and then the client send verified certificate back [8].

EAP-TTLS a PEAP

Both TTLS and PEAP were developed in response to the PKI barrier in EAP-TLS. The structure of TTLS and PEAP are quite similar. Both are two-stage protocols that establish security in stage one and then exchange authentication in stage two. Stage one of both protocols establishes a TLS tunnel and authenticates the authentication server to the client with a certificate. (TTLS and PEAP still use certificates to authenticate the wireless network to the user, but only a few certificates will be required, so it is much more manageable.) Once that secure channel has been established, client authentication credentials are exchanged in the second stage [7].

EAP-FAST

The EAP-FAST (Flexible Authentication via Secure Tunneling) is a protocol proposal by Cisco Systems as a replacement for LEAP. The protocol was designed to address the weaknesses of LEAP while preserving the "lightweight" implementation. Use of server certificates is optional here. It uses a Protected Access Credential (PAC) to establish a TLS tunnel in which client credentials are verified. EAP-FAST has three phases. Phase 0 is an optional phase in which the PAC can be provisioned manually or dynamically, but is outside the scope of EAP-FAST. PAC provisioning is still officially Work-in-progress, even though there are many implementations. PAC provisioning typically only needs to be done once for a RADIUS server, client pair. In Phase 1, the client and the AAA server uses the PAC to establish TLS tunnel. In Phase 2, the client credentials are exchanged inside the encrypted tunnel [5].

III. DESCRIPTION OF PROPOSED COMMUNICATION

Proposed system has to provide:
  (1)   establishing a connection and procedure for session negotiation,
  (2)   device authentication procedure,
  (3)   user authentication procedure,
  (4)   procedure for generating a key set for data encryption and data integrity check, updating a session and storing the ky set,

Procedures (1) a (2) will be describe detailed in sequence with Fig.1, procedure (3) in sequence with Fig. 2.

Referring to Fig.1, each of BSs establishes radio connection with Mobile Station (MS) within its service area, i.e. cell, and communicates with the MSs via the radio connections. When MS is in idle mode, SRNC is controlling signaling of the MS through the BS. The BS connects the MS to a packet data network such as the Internet through an Access Gateway (AG). In Fig.1, significant network entities in the packet data network, i.e. Home Agent (HA) and an Authentication, Authorization and Accounting (AAA) server, are shown. If authenticator for device authentication of the MS is in the SRNC, the SRNC having an interface with the AAA sever is used for device authentication, as will be described below.

Interfaces exist between the BS and the SRNC and between the AG and the SRNC for managing the mobility of the MS, and a data path exist between the AG and the BS. To authenticate the MS, the SRNC is provided with a device authenticator (not shown) and the AG has a user authenticator (not shown). For this time it could be described that the AG and the SRNC are incorporated into a single physical entity for authentication, the same effect is achieved as far as an appropriate interface between the AG and the SRNC, even when the SRNC is configured to be a separate physical entity.
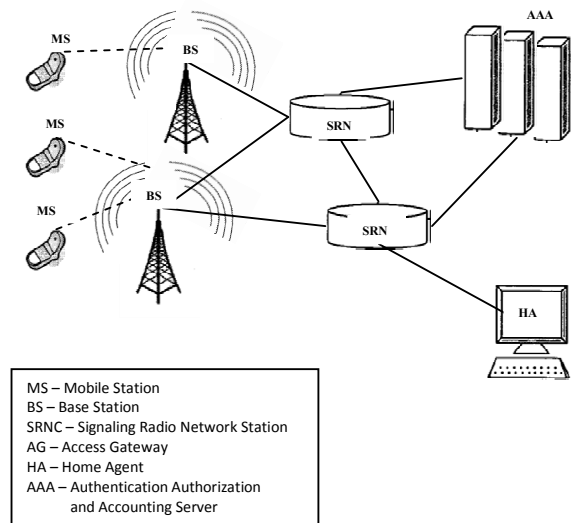
Fig. 1. diagram of mobile communication network environment

Upon received of a connection request message (Fig.2), for example, a route request message from the MS (Connection Request), the BS transmits a Context Request message to the SRNC to request session. When the SRNC delivers a Context Response message including session information to the BS, the BS transmits to the MS a connection response message for the connection request of the MS. Then a session negotiation/configuration is carried out.

After the connection setup and session negotiation is completed, device authentication is performed in next steps. The BS transmits an AR EAP Start message to the SRNC to trigger the SRNC`s transmission of an EAP Request message. When the SRNC transmits an AR EAP Payload message with

an EAP Request/ID encapsulated to the BS, the BS transmits an EAP Request/ID Radio Link Protocol (RLP) message including the EAP Request/ID to the MS by the RLP.

The MS transmits an EAP Response/ID RLP message having an EAP Response/ID with NAI to the BS by the RLP in response to the Request/ID and the BS transmits an AR EAP Payload message having the EAP Response/ID encapsulated to the SRNC. In next step, the SRNC transmits to the AAA server the EAP Response/ID in an AAA Access Accept message such as a Remote Authentication Dai-In User Service (RADIUS) access request message or an access request message based on the AAA protocol. Thus, an EAP negotiation is made between the MS and the AAA server according to an EAP method through SRNC.

When the EAP negotiation is completed, the AAA server transmits an EAP Success message and D-MSK to the SRNC by an AAA Access Accept message. It can be further contemplated as another exemplary embodiment of the present invention that the SRNC generates a PMK using the D-MSK and stores the PMK. In last step, the SRNC transmits an AR EAP Payload message with the EAP Success message encapsulated to the BS. The EAP Success message is delivered from the BS to the MS by an RLP message.

Referring to second part of Fig. 2 (red highlighted), the BS transmits an AR EAP Start message to the AG to trigger the AG`s transmission of an EAP Request message. Alternatively, the SRNC may received the AR EAP Start message from the BS and relay it to the AG. The AG transmits an AR EAP payload message with an EAP Request/ID encapsulated to the BS. Alternatively, the SRNC receives the AR EAP Payload message from the AG and then relays it to the BS. The Bs transmits an EAP Request/ID RLP message including the EAP Request.ID to the MS by the RLP [9].

The MS transmits an EAP Response/ID RLP message having an EAP Response/ID with a NAI to the BS by the RLP in response to the EAP Request/ID and the BS transmits an AR EAP Payload message having the EAP Response/ID with the NAI encapsulated to the SRNC. SRNC relays the AR EAP Payload having the EAP Response/ID with NAI encapsulated to the AG. If the AG and SRNC are configured to be a single physical entity, these steps could be just one process that takes place in an internal interface of the physical entity. In next step the AG transmits to the AAA server the EAP Response/ID with NAI encapsulated in an AAA Access Accept message such as a RADIUS access request message based on Diameter AAA protocol. Thus, an EAP negotiation is made between the MS and the AAA server according to an EAP method through the SRNC and AG.

When the EAP negotiation is completed, the AAA server transmits an EAP Success message and a U-MSK to the AG by an AAA Access Accept message. Although both or either of the user authentication and the device authentication can be performed according to the service provider`s choice, this invention uses a key used for the user authentication as a root key for a subsequent procedure if the user authentication follows the device authentication. Therefore, the AG induces an R-MSK from the U-MSK received form AAA server and transmits the R-MSK to the SRNC by an AR EAP Payload message with the EAP Success message encapsulated.

In next step, the SRNC relays the AR EAP Payload

message with the EAP Success message and the R-MSK encapsulated to the BS. If a PMK is generated using the U-MSK, the SRNC induces the PMK from the R-MSK generated from the U-MSK and transmits the EAP Success message to the MS by RLP message. The MS stores the PMK acquired from the EAP Success message from use during a session [7].
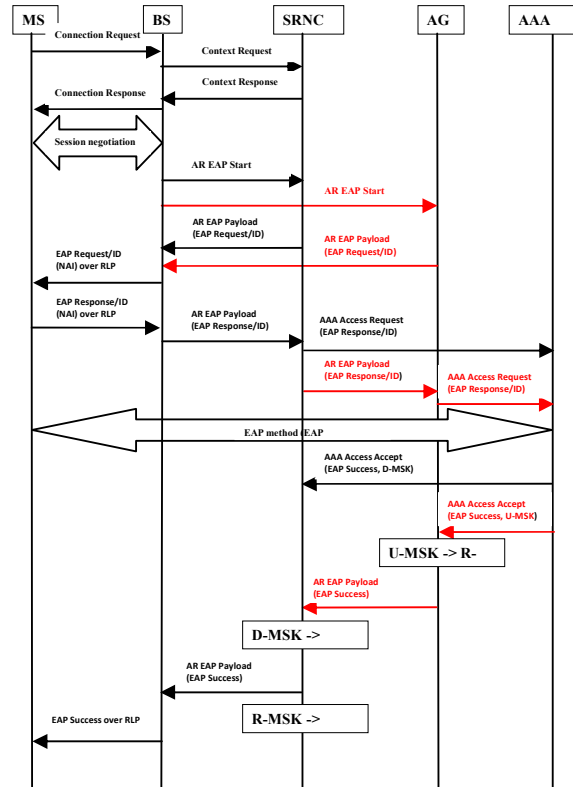
Fig.2.: *data flow diagram for connection and authentication process and for user authentication*

IV. THE SRNC FUNCTIONALITY DESRIPTION IN DEFINED NETWORK ENVIRONMENT

The SRNC receives a request message including a Context Request or Session Fetch Request from the BS and replies with a Context Response message or a Session Fetch Response message.

The Context Request message and the Session Fetch Request message request a context including session information to estblish a communication path. Upon receipt of an Authentication Relay (AR) EAP Start message-requesting authentication from the BS, the SRNC transmits to the BS an AR EAP Payload message in which an EAP Request message with Network Access Identifier (NAI) is encapsulated. The EAP Request message with the NAI is referred to as an EAP Request/ID, where the ID is an identifier field in the EAP Request message.

Upon receipt of an AR EAP Payload message having an EAP Response message with the NAI encapsulated from the BS, the SRNC transmits an AAA Access Request message with the EAP Response message to the AAA server in order to perform an EAP authentication procedure. In next steps, the EAP Response message with the NAI is referred to as the EAP Response/ID. In next step an EAP conversation can be

conducted between the SRNC and the AAA server according to an EAP method. When the EAP conversation is completed, the SRNC receives an EAP Success message indicating success of the EAP authentication and a Device-Master Session Key (D-MSK) related to device authentication from the AAA server by an AAA Access Accept message. The EAP Success message and the D-MSK are encapsulated in the AAA Access Accept message. The D-MSK is a master key for use in device authentication. The SRNC can use the D-MSK for generating a Pairwise Master Key (PMK).

If the SRNC has not received the AR EAP Start message directly from the BS, all process continues by receiving an AR EAP Payload message having an EAP Response/ID with NAI from the BS and the SRNC relays the AR EAP Payload message having the EAP Response/ID to the AG. The EAP Response/ID is generated by the BS in response to an EAP Response/ID received directly from the AG.

In next step, the SRNC receives from the AG an EAP Payload message in which an EAP Success message related to the EAP Response/ID is encapsulated. This EAP payload message with the EAP Success message encapsulated includes a Root-MSK (R-MSK) that the AG has generated based on User-MSK (U-MSK). The U-MSK is used for user authentication. The AAA server generates the U-MSK based on a long-term credential and provides it to the AG, for generation of the R-MSK. The R-MSK can be used for generating a PMK. The SRNC generates the PMK using the R-MSK. The PMK is used for generation of data encryption relayed keys to be used during a session, for generation of keys to be used for data integrity check or for generation of a Session Root Key (SRK). I another example, the SRK can be used for generation of data encryption-relayed keys to be used during a session or for generation of data integration keys.

The D-MSK and U-MSK are used for device authentication and user authentication respectively. The D-MSK and U-MSK are induced from the long-term credentials or an Extended-MSK (E-MSK) induced from these terms. If the U-MSK is used, the SRNC receives the R-MSK generated from the U-MSK from the AG, generates the PMK using the R-MSK and generates the data encryption-related keys using the PMK.

V. SUMMARY

A method for performing device authentication and user authentication of a Mobile Station (MS) in a mobile communication network, comprising the steps of:

Establishing a connection between the MS and SRNC that controls communications of the MS through a BS;

Receiving a Device-Master Session Key (D-MSK) for device authentication of the MS from an Authentication, Authorization and Accounting (AAA) server thathas completed Extensible Authentication Protocol (EAP) negotiation with the MS and storing the D-MSK by the SRNC, when the BS triggers an EAP authentication after the connection estblishment;

Recaiving a Root-MSK (R-MSK) from an Access Gateway (AG) and storing the R-MSK by the SRNC after the connection establishment, the R-MSK being generated using a User-MSK (U-MSK) for the user authentication of th MS received from the AAA server by the AG;

Generating a Pairwise Master Key (PMK) for use during a session using at least one of the D-MSK and the R-MSK by the SRNC; and

Generating a key set using the PMK by one of the BS and the SRNC, for use in at least one of data encryption, data integrity check and session management during he session.

VI. CONCLUSION

Authentication and security method for a mobile communication network according to the present proposal is also applicable to other mobile communication systems having a similar technological and channel structure with jus slight modifications made within a scope of present proposal.

REFERENCES

[1] Interoperability Specification (IOS) for Ultra Mobile Broadband (UMB) Radio Access Network Interfaces, V&V Version, www.3gpp2.org, July 2007
[2] Simple IP Services for Converged Access Network, www.3gpp2.org ballot version, July 30, 2007
[3] EAP Method for 3rd Generation Authentication and Key Agreement (EAP-AKA) IETF RFC 4187, January 2006
[4] Extensible Authentication Protocol (EAP)", RFC 3748, June 2004
[5] EAP Method for 3rd Generation Authentication and Key Agreement (EAP-AKA) IETF RFC 4187, January 2006
[6] Remote Authentication Dial In User Service (RADIUS), IETF RFC 2138, April 1997
[7] PPP EAP TLS Authentication Protocol, IETF RFC 2716, October 1999.
[8] EAP Tunneled TLS Authentication Protocol (EAP-TTLS), IETF Internet Draft, February 2002.
[9] EAP Method for GSM Subscriber Identity Modules (EAP-SIM), IETF

Automatic segmentation of moving objects on real conditions

[1]*Vladimír Jeleň*

[1]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]vladimir.jelen@tuke.sk

*Abstract –***It is important to monitor an object and evaluate its characteristics such as speed and direction of its movement, size, shape, etc in many areas of science. In our case it was necessary to analyze the movement of mini-pig infected by neurodegenerative disease. As it was not necessary to use the real time analyzing, we used the offline method of analyzing, what allowed us to use more time-consuming methods which afforded more accurately analysis.**

Keywords – motion tracking, segmentation, grab-cuts, threshold, background, seed region, bi-elliptical model, kalman filter

I. INTRODUCTION

In recent time, it is still more popular to use the knowledge from technical science also in medicine or physics, etc. Our algorithm was created as a reaction to necessity of analyzing the influence of neurodegenerative diseases to behavior of living organisms. In this article, I tried to summarize our approach to the task and to explain the algorithm we used.

II. MATERIALS AND METHODS

A. Hardware

We used computer with 2GHz Dual Core CPU, 3GB of RAM and 64MB VGA+ for our segmentation method. It is a common configuration.

B. Software

We used the OpenCV and Visual studio 2005 for creation of all needed applications. The applications were developed in C++ language.

III. IMAGE ANALYSIS AND MOTION TRACKING

First of all it is necessary to divide the video into individual frames for motion trFacking of an object. We found out, that if the video was encoded by DivX, it is difficult to extract individual frames. That is why we encoded each analyzed video in mjpeg format. The video was mostly framed with frequency 25 frames per second. In our case, we used frequency 1 frame per second for motion tracking. We

extracted the frames from video by the application which was developed exactly for this project. It is called *VideoDecompiler*. You can see it in the figure 1. The application allows selecting of video for decompilation and frequency of framing. Individual frames are saved into dedicated directory.

Fig. 1 video decompiler

Creating of background

For time-effective detection of object it is adequate to eliminate all unnecessary information from each frame. We achieved it so we deducted background from each frame. There remained only object and noise in the analyzed image.

The background we can get by two methods:

1) We recorded the background first we placed the object in the monitored field.

2) We selected stack of frames, so we removed the object from the original place to another. In the last frame the object did not cover any pixel, which was covered in the first frame. The analyzed object was lighter than background so we found the minimum of every pixel in every frame in the stack.

Fig. 2 Analysed frame

Fig. 3 Background

Deducting of background from frame

As it was mentioned, for improving time-effectiveness of image processing it is adequate to eliminate all unnecessary information. We deducted background from frame and we got more homogenetic image. In this image we applied binary threshold so we eliminated most of noise. Even though little of noise still remained there. From this moment we worked just with the binary image. Other smoothing of image we did with morphological operation *opening*.

Fig. 4 Binary threshold

Opening – for eliminating small objects

We used operation *opening* for eliminating little of noise from image (fig. 4), so we smoothed object and eliminated small areas caused by noise. It was necessary for saving time

for other image processing. Usually the image after the operation is almost perfect smooth and contains only binary information about the object. The image may contain some false objects. It is caused by imperfection of operation *threshold*. We eliminated such like objects by analyzing of property of objects of the image, specifically their content.

Eliminating all objects based on their content

We knew what the object should contain, that is why we filtered off all objects which did not have the adequate content. All objects which contained more or less pixels as we needed were eliminated. Objects with less than 500 pixels and more than 4000 pixels were eliminated. In our case the objects contained about 2000 pixels. Frames with resolution 352x288.

Finding the center of gravity

Usually the object after this operation was equally detected. We found the center of gravity of the object in binary image which contained information of object's position and content. It is showed in the figure 5.

Fig. 5 Object and center of gravity

Finding the farthest points and creating lines with the middles in the center of gravity

The object which we were analyzing in our experiment was mini-pig. Its basic property was its oval shape. Its shape could be more complicated than just oval. We described it in the article [5]. According to this knowledge we found the two farthest points of the object which were crossing the center of gravity.

Fig. 6 Angle-sceleton of minipig

Creating seed region

According to the center of gravity and two marginal points it is possible to create a curved line – *angle-skeleton* which describes how is the pig flexed, where is it looking and how is it removed in the field. We linked two ellipses to the angle-skeleton. The ellipses had one common point in the pig's center of gravity. We called this object *bi-elliptic model* [8]. We applied this model as a seed region in the *grab-cut* method.
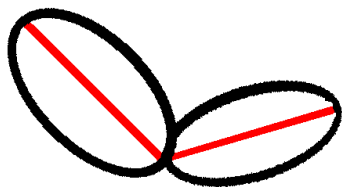
Fig. 7 Bi-Eliptic model

Applying grab-cut

We used the grab-cut method for finding the best edge of object. We used the method also in [9]. We chose this method because it was adequate for finding edges of objects also in the complicated images. This method in the input needs analyzed image, region of pixels which absolutely belong to the object and region of pixels which absolutely do not belong to the object.

We generated the input parameters according to our bi-elliptic model. We reduced the object represented by our model to 60% and we used the reduced object as a seed region, the region which absolutely belonged to the object. Thereafter we enhanced the bi-elliptic model to 200% and edge of the model represented the region which absolutely did not belong to the object and at the same time allocated the region behind which we would not apply the grab-cut. The edge was coveted in the accrued area.

Fig. 8 Bi-Eliptic model on real condition

Editing kalman filter and predicting next center of gravity

We also used *Kalman filter* for further improving motion tracking of object. It allowed us to predict position of next center of gravity according to all previous. The improving of algorithm also serve to accelerating of analyze and improving the accuracy.

Practically it worked as follows:

1. Setting the initial attributes of Kalman filter.
2. Predicted attributes served to improve the filtering the object from noise.
3. After final recovery of the real edge of object, the center of gravity was calculated and kalman filter was updated.

IV. EXPERIMENTAL RESULTS

We found out, that our algorithm was stabile and was adequate for conditions with stabile lighting. In conditions with unstable lighting, it was necessary to calibrate the algorithm and edit the way we got the information about background. The algorithm was stabile also in case when the monitored object had very similar properties as the background.

V. CONCLUSION

Our approach was very effective from sight of finding the properties of object, such as shape, position, rotation, etc. We improve the effectiveness of the algorithm by combination of modern and classic methods. This method is adequate for motion analyses of different objects in different conditions.

REFERENCES

[1] GREIG, D. – PORTEOUS, B.SEHEULT, A.: Exact Maximum A Posteriori Estimation for Binary Images, J. Royal Statistical Soc. – Series B, vol. 51, no. 2, pp. 271-279, 1989

[2] ISHIKAWA, H. – GEIGER, D.: Segmentation by Grouping Junctions, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 125-131, 1998

[3] BOYKOV , Y. – KOLMOGOROV, V.: Computing Geodesics and Minimal Surfaces via Graph Cuts, Proc. European Conf. Computer Vision, pp. 232-248, 1998

[4] LEGAULT, R. – SUEN, CH.: Optimal Local Weighted Averaging Methods in Contour Smoothing, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 19, no. 8, august 1997

[5] TOMORI, Z. – JELEŇ, V. – JANÁČEK, J.: Animal motion tracking using advanced image segmentation methods, 6th International Conference: Structure and Stability of Biomacromolecules, Košice, Slovakia, 2009, pp. 69-70.

[6] P. Das, *et al.*, "Semiautomatic segmentation with compact shape prior," *Image and Vision Computing,* vol. 27, pp. 206-219, 2009.

[7] C. Liu, *et al.*, "Interactive Image Segmentation Based on Hierarchical Graph-Cut Optimization with Generic Shape Prior," Image Analysis and Recognition, Proceedings, vol. 5627, pp. 201-210, 957, 2009.

[8] JELEŇ, V. – JANÁČEK, J.– TOMORI, Z.: Mobility Tracking by Interactive Graph-Cut Segmentation with Bi-elliptical Shape Prior, The 8th International Symposium on Applied Machine Intelligence and Informatics. Herľany, Slovakia.

[9] JELEŇ, V. – DEMJÉN, E. – TOMORI, Z.: Segmentation of human head contour by graph-cuts algorithm and its use in biometric security systems, The Fourth Biomedical Engineering Conference of Young Biomedical Engineers and Researchers YBERC 2010. . Kosice, Slovakia.

Automatic Subgoals Discovering Method In Reinforcement Learning Using Newness In State Space

^{1}*Marek Lapko,*^{2}*Pavol Maliňák*

^{1,2}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

^{1}marek.lapko@tuke.sk, ^{2}pavol.malinak@tuke.sk

Abstract—In this paper, we present an algorithm that automatically discovers subgoals in reinforcement learning. In the proposed algorithm, the subgoal is a narrow part of a state space - "bottle neck" that lies between (or connects) two regions. The main idea of the algorithm is that, as an agent goes through the subgoal and enters the new region the agent starts exploring the part of the state space in which it has not been before for a specific time. The aim of discovering subgoals is to speed up learning by allowing transfer of knowledge on various levels of abstraction. The performance of our algorithm is illustrated in so-called rooms navigation problem.

Keywords—automatic subgoal discovering, option, Q-learning

I. Introduction

Reinforcement learning has proven to be a powerful tool in tasks, where (almost) all information we can provide learning agent is a goal to reach. The goal is determined by a reward, usually at the end of the episode. The aim of the agent is to learn by trial and error to get the highest reward along the path and thus reach the goal [1]. As the agent is provided just with a little information, it takes the agent a long time to learn, especially in the case when the agent is a real robot, not a simulation. One way, how we could speed up the learning, is to reuse the knowledge from previously learned tasks. To transfer the knowledge, we consider hierarchical reinforcement learning approach. The knowledge transfer in various levels of abstraction allows us to use even parts of knowledge in a new task.

In this paper, we focus on the problem, how to find subgoals in a task. These subgoals divide the task into the parts, subtasks. Every such a task could be assigned to an option [2], which we can see as an action on the higher level of abstraction and which is easy to transfer [3][4].

The subgoals, we can see as a "funnels" or "bottle necks" in the state space [5]. In general, the bottle neck is a state (or a group of states), that lie on the path to the goal and it is critical to go through this state. There are several methods for finding such subgoals. In [6] as subgoals there are determined the states with high frequency of visit or the states with a high reward gradient. Other approach to discover subgoals automatically is to use diverse density [5]. Krechmar et al. in [7] proposed algorithm that uses a combination of a frequency and a distance of a state from the start or from the end state. Subgoal discovered by algorithm in [8] is a state with a high gradient of a number of states leading to the state. Simsek

in [9] considered similar idea to ours and defines a relative novelty of a state in a transition sequence.

II. Reinforcement learning

Reinforcement learning is a kind of learning, in which an agent, by interaction with an environment, tries to reach a goal. Usually, this problem is described as Markov Decision Process (MDP) [1]. An MDP is given by the tuple $\langle S,A,T,R \rangle$, where S is a set of states, that the agent can be in, A is a set of available actions, $T : S \times A \times S \rightarrow [0,1]$ is a transition function, which defines the probability of transition from the state $s \in S$ to the state $s' \in S$ when choosing action $a \in A$ in state s. $R : S \times A \rightarrow \mathcal{R}$ defines a reward, that the agent receives from the environment when choosing action a in state s. The goal is to find an optimal policy $\pi : S \times A \rightarrow [0,1]$.

Q-learning is one of the most used off-policy RL algorithms. The update rule for Q-learning is defined as:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \\ + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \tag{1}$$

where $Q(s_t, a_t)$ is the action-value function, which is a value of choosing action a in state s and following policy π and r_{t+1} is a actual reward given by the environment in time $t+1$.

As our goal is to reuse the knowledge on various levels of abstraction, subgoal is represented by an option framework. Option is extended action - closed-looped policy for taking action over period of time [2]. Option is defined as a triple $\langle I, \pi, \beta \rangle$, where $I \subseteq S$, is an initiation set of states, states in which an option is available. π is the option's policy and $\beta(s)$ is the probability, that option will terminate in state s.

III. Newness in state space

As we have already mentioned, the subgoal, our algorithm is able to discover, is defined as a state, which connects two state regions. By region, we understand set of states that are mutually easily accessible, which means, that there are more different (possible) trajectories in the state space between each two states. Between two regions, there are only few (or even one) possible trajectories that connect states from two different regions. On these trajectories lie states, that we consider to be subgoals.
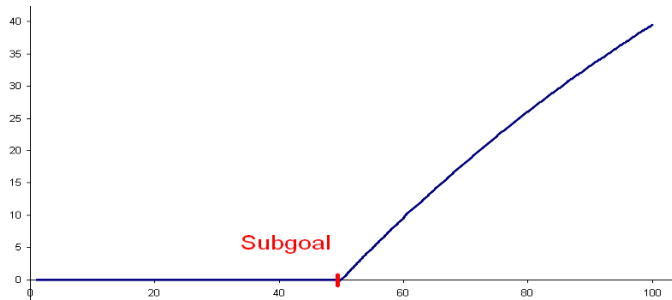
Fig. 1. Subgoal template. The subgoal template in a time series of the state newness for parameter $\beta = 0.99$. The red bar indicates position of the subgoal. The x-axis refers to t and the y-axis refers to N according to 2.

Fig. 2. Two-room gridworld environment. There is an agent (robot) with key in the upper left corner. The key position represents a yellow square. The door is a gap between two rectangular obstacles in the middle of the environment. In the upper right corner, there is a goal (red square) that the agent tries to reach.

One of the main features of the reinforcement learning algorithms is a *trial and error* approach. It means that there is a certain probability of choosing a random action[1] and the agent, in order to find the optimal policy, is thus exploring the state space.

Our algorithm is based on the idea that as the agent is exploring the state space, more states are being visited and the region in the state space becomes familiar for the agent. After several time steps (depending on a size and a discretization of the state space) agent occurs at states that has already been visited before. When the agent crosses some potential subgoal and enters a new region it starts occurring at unvisited states. So in the proposed algorithm we are looking for the states in which there is a significant change in a temporal presence of the visited and unvisited states.

The proposed algorithm runs in parellel with learning algorithm (Q-learning). During learning we just need to store history of the visited states s_t for every time step of the episode and the information of the newness N_t of a current state s_t, where

$$N_t = \begin{cases} \beta N_{t-1} + 1 & \text{if } s_t \text{ is visited for the first time} \\ \beta N_{t-1} & \text{otherwise.} \end{cases} \quad (2)$$

The β parameter is from $[0, 1]$.

The next step is to find a position (t) in the history of the newness where the value changes more significantly. According to (2), the history of N before and after the subgoal should look like on Fig. 1, where the value of N_t is not increasing in the current region until it enters a new region. To find such a template τ in the given history, we used normalized cross-correlation (3), where template τ is shifted by u steps in x direction [10].

$$r = \frac{\sum_x (f(x) - \bar{f}_u)(\tau(x - u) - \bar{\tau})}{\sqrt{\sum_x (f(x) - \bar{f}_u)^2 \sum_x (\tau(x - u) - \bar{\tau})^2}}, \quad (3)$$

where $\bar{f}_u$ denotes the mean of $f(x)$ within the area of the template τ shifted by u steps and $\bar{\tau}$ is the mean value of the template τ.

After every episode, a candidate for the subgoal is the state with local maxima in the series of correlation coefficients higher than a specified threshold. The position of the potential subgoal is shifted within the template by n steps, where n is the position in the template, where N starts to increase.

[1]e.g. ϵ-greedy policy
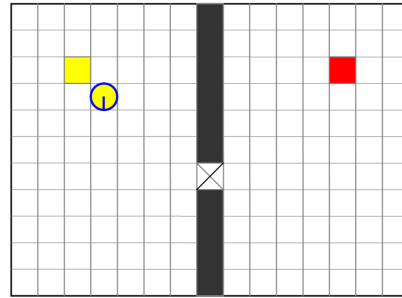
IV. EXPERIMENTAL RESULTS

All experiments were conducted in a simulated environment. The environment's properties were set in the way to be similar to our LEGO robot environment.

Our environment has a planar rectangular shape, with two obstacles placed to create a "door" and to separate two regions (Fig. 2). The environment is designed the way there two obvious subgoals - the key (to open the door) and the door. Without the key, the agent can not go through the door.

We used Q-learning algorithm to find optimal policy with a table representation of knowledge and therefore we had to discretize the state space and the action space. The state space was represented by agent's position in the environment, where the position was described by the x and y position and the agent's orientation (an angle with the x axis) and the information whether the agent has the key (0;1). We had three discrete actions: turn right $(+\alpha)$, turn left $(-\alpha)$ and go straight.

A. Q-learning

The size of the environment was 600×500 with discretization step 20 units in both dimensions. The orientation of the agent was discretized by $\frac{\pi}{4}$ step. The result was, that we had only eight possible orientations. There were three possible actions for the agent turn $\pm\frac{\pi}{4}$ and go straight 20 units. Q-learning parameters were set to the values: $\gamma = 0.9$, $\epsilon \in [0.05, 0.01]$ and $\alpha \in [0.1, 0.03]$. The agent got reward $+1.0$ when it reached the goal, otherwise it was punished every step with punishment -0.01.

B. Newness in state space

The algorithm parameter β was set to 0.99. The template had length set to 100 steps, with the subgoal position at 50. Values of the template before the subgoal position were set to 0.0 and were increasing according to (2) after the subgoal position.

After every episode, we selected the potential subgoals with local maxima in the series of the correlation coeffitients higher than 0.98.

In Fig. 3, we can see how many each state has been selected as a potential subgoal after first 100 episodes. When finding subgoals in this task, we can omit the orientation of the agent, as for the subgoal the important is just the position not the orientation of the agent [11]. With this assumption, the count of every state $withkey$ is depicted in Fig. 4 (a). The key position has been selected 9 times and the door position has
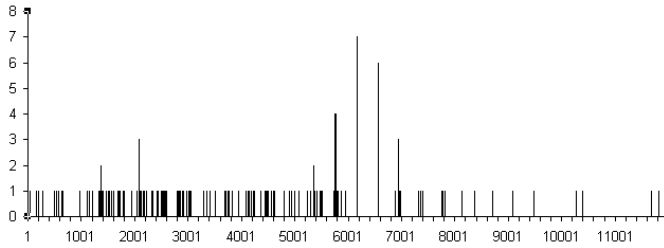
Fig. 3. Count of potential subgoals. The x-axis represents state index and the y-axis represents the number of the state selection as a subgoal.

a.)

b.)

c.)
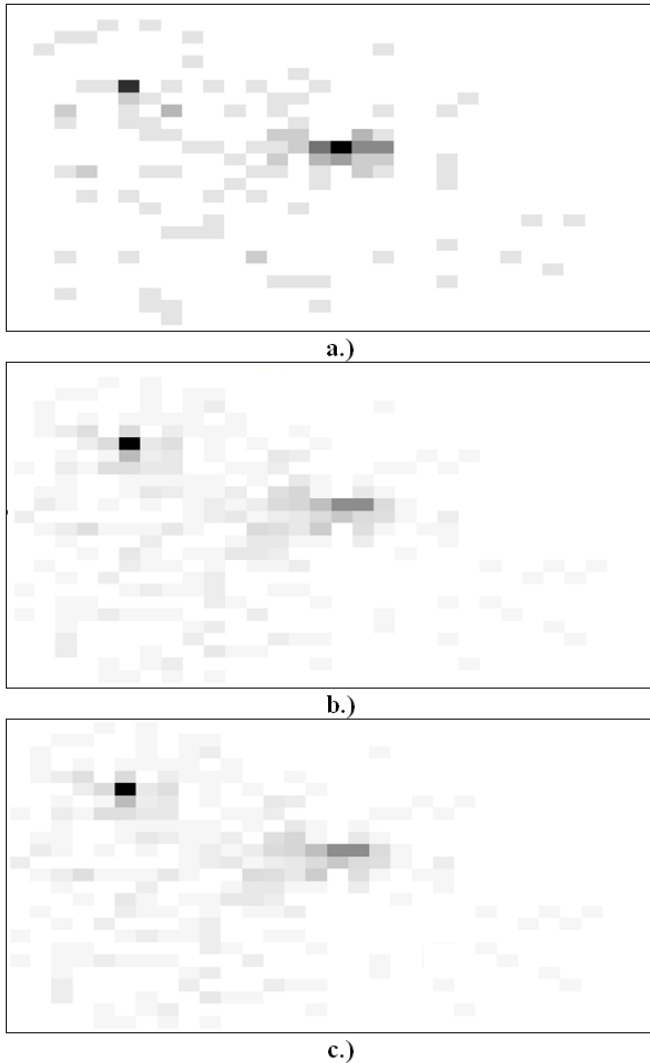
Fig. 4. State visitations histograms. Black color represents maximal number in each histogram. (a) depicts situations after 100 episodes, (b) after 500 and (c) after 1100

been selected 11 times. After first 500 episodes (Fig 4(b)), the state at the key position has been selected 43 times, whereas the state at the door position only 19 times. After another 600 episodes (Fig. 4 (c)), the count of selections remains the same.

According to conducted experiments, the subgoals are more obvious after the first 100 episodes when we can distinguish them easily. Later on, the door-subgoal, which is closer to the goal, becomes less significant in contrast to the key-subgoal but is still visible. We can see a gradient in the histogram with a local maximum in the position of the door-subgoal.

The reason is learning of the agent. The policy of the agent selects the action according to $Q-table$ with the highest value.

As the agent is rewarded at the very end, the reward "spreads" from the goal, back to starting positions. Therefore the agent is doing less exploration of the environment around the subgoals that are closer to the goal. The decrease in exploration results in the situation when the difference between regions in the number of the visitation of states is not so significant.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented the algorithm capable of discovering all subgoals in a simple gridworld environment within reinforcement learning. To discover the subgoals, the algorithm utilizes the difference in the property of states before and after them, which we call *newness*.

In the future work, we would like to focus on finding subgoals on real domains where subgoals do not have to be so evident.

REFERENCES

[1] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning)*. The MIT Press, March 1998.
[2] R. Sutton, D. Precup, and S. Singh, "Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning," *Artificial intelligence*, vol. 112, no. 1, pp. 181–211, 1999.
[3] M. Lapko and R. Jaksa, "Reuse of knowledge within reinforcement learning in mobile robotics," *13th International Conference on Cognitive and Neural Systems 2009 (accepted)*, 2009.
[4] M. Lapko, "Reuse of knowledge in reinforcement learning (in Slovak)," *Thesis proposal*.
[5] A. McGovern and A. Barto, "Automatic discovery of subgoals in reinforcement learning using diverse density," pp. 361–368, 2001.
[6] B. Digney, "Learning hierarchical control structures for multiple tasks and changing environments," p. 321, 1998.
[7] R. Kretchmar, T. Feil, and R. Bansal, "Improved automatic discovery of subgoals for options in hierarchical reinforcement learning," *Journal of Computer Science and Technology*, vol. 3, no. 2, pp. 9–14, 2003.
[8] S. Goel and M. Huber, "Subgoal Discovery for Hiearchical Reinforcement Learning Using Learned Policies," 2003.
[9] O. Şimşek, "BEHAVIORAL BUILDING BLOCKS FOR AUTONOMOUS AGENTS: DESCRIPTION, IDENTIFICATION, AND LEARNING," 2008.
[10] K. Briechle and U. Hanebeck, "Template matching using fast normalized cross correlation," vol. 4387, pp. 95–102, 2001.
[11] M. Lapko, R. Jaksa, and P. Sincak, "Automatic subgoals discovering method with state abstraction in reinforcement learning," *Multidisciplinary Symposium on Reinforcement Learning*, 2009. [Online]. Available: http://msrl09.rl-community.org/abstracts

Basis for assisted software language creation

[1]*Dominik LAKATOŠ*

[1]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]dominik.lakatos@tuke.sk

Abstract— **These paper describes fundamental needs for creation of new software languages and provides overview of domain specific languages. It analyzes problems connected with creation of domain specific languages. Main goal is to describe common language model usable as basis for generating language specific tools. Model can also be used for language composition. Then we discuss issues concerning possible tools like parsing, printing. It presents possible input forms for language specification for common language model. By providing assisted language creation we can lower general cost of language.**

Keywords— **computer languages, composition, domain-specific languages, model, printing, tools**

I. INTRODUCTION

Every time we look around us, we get some information from our surrounding. In order to understand such information we need to have knowledge of its meaning, its semantic. In other words, we have to know appropriate language. Almost everything can be described as some kind of language. We have traditional languages for communication, in oral or written form, but also our gestures transmit plenty of information for other people. Same it can be for communication between animals, or even art in any form can propagate some idea. Most of these so-called languages exist only in abstract form; usually we even cannot explain exactly the meaning. In this times we live our lives interconnected with machines, computers. They help us in different ways. But it is still very hard to make computer easily understand some new language, to teach them new language. In order to do so, we need to understand grammar theory, parsing, definition of semantics and have plenty of additional skills.

In order to create something new with computer, we need to understand some programming language. This kind of language is hard to understand for ordinary people, or even for people with some specific knowledge. Imagine security expert trying to set non standard rules for securing building or even computer network, usually you have to understand existing possibilities and work only with those, it is not possible to write something new, because it would need a lot of work from someone else, or it would be even impossible to add to existing system. So our security expert has to spend his time thinking about how it can be done, instead of thinking about solution for next issue.

II. HELP OF DOMAIN SPECIFIC LANGUAGES

Let's try to transform our problem to software language territory. Usual way for developing computer systems is through usage of general purpose languages like C, C++, Java, C# and many more. To be productive and so called "good programmer" it is not sufficient to just write programs solving specific problems, but you have to do language design as well [1]. Creation of new language can be expensive task, but it can help us in many different ways. We can be able to let our domain expert (for example security expert) to write his problems in words from his field of expertise and read outputs from computer in this same language constructs. So we would use domain specific terms which are understandable also for users and not only for developers. Designed language does not have to be extremely complex and universal as are general purpose languages.

Domain specific languages promise advantages over general purpose languages. Anyone who can write same thing in less word and simpler form would be more productive, because as mentioned in [2], programmers tend to produce statements at the same rate whatever language they use. Only a few programmers go to trouble of creating new language as it is a problematic area for no skilled person and it can lengthen development process. Even if you create new language, it can be hard to bring to practical usage as it is new syntax to learn and it can be expensive to maintain [3]. This is why we usually can see domain specific terms only used in general purpose languages, which does not support abstraction very well. You ordinarily end up with code containing a lot of computer specific programming noise and lot less written code is really useful for solution of domain problems. So as you can see problem in developing new language is mainly in it cost and lack of experience to do so.

Research to find techniques for developing secure, reliable and error free software is a long lasting process [4]. We can choose from different proposed engineering methods for such development: from waterfall model [5], to aspect oriented frameworks [6]. We would like to aim to language-based software engineering frameworks, which allows for creation of systems with use of own special languages, mostly domain-specific languages [3]. This way we can also use domain experts as application developers to describe anything needed for created new system.

III. ISSUES OF LANGUAGES

We can say that the ultimate issue for not creating and using new languages is price for its development and maintenance. We can lower this price with tools supporting and helping in process of creation of new software language. Such tool can help overcome problems facing parsing issues for classical

textual languages. It can help you choose proper parsing type or provide tips for syntax. It can generate tools for serialization of internal sentence in memory to its textual representation (we call this printing). Creating such supporting tools should be main purpose in providing more accessible language engineering.

Even if we have some domain specific languages created and used, there is another problem emerging. Have you ever tried to use domain specific language to configure behavior of program written in general purpose language or even use just XML for the same thing? It can be very handy, as long as you don't need to change name of something (e.g. rename class). In this instance you usually have to change this name manually also in your file with DSL or XML, even if in program you can use refactoring. There is no standard way to describe different languages and composition of such languages.

We would like to try to provide our view to this problem and possible solutions in this paper. It will be also presented on some of our tools. The main issue is to create common language model. Such model can serve as central part for describing each software language and as starting point for creation of supporting tools. These tools help to use the new language right away without need to create them specifically for each language. They can provide automatic textual output, language structure traversing, syntactical IDE highlight support and many more. Common language model can also help in creating composition of different languages and describing overlapping concepts.

IV. LANGUAGE MODEL DESIGN

Original framework for modeling language, which is our base framework, was completely described in [7]. It was based on writing of language implementation in annotated Java classes. Information provided in classes was used directly to generate language processor (parser), for designed language. This design was susceptible to unsupported constructions due to robust possibilities of Java classes. When we tried to create new tools, we had to deal with all information provided in Java classes again and again. So we decided to create our own internal model for description of language. Such model can also help with language composition. We used traditional language names for naming parts in model. For example, main used concepts are:

- language,
- concept,
- abstract syntax,
- concrete syntax,
- patterns

Simple connection between concepts of common language model is shown on Fig. 1. Each language consists of main concept named language, which can have one or more concepts. Concept can be divided to description of abstract syntax and concrete syntax. Abstract syntax can be further divided to properties of concept. Similarly, each context's concrete syntax consists of none, one or few notations.

We should be able to describe our language model concepts using its own structure, as it is done e.g. in UML [8]. This way it is possible to create tools also for our common language model.
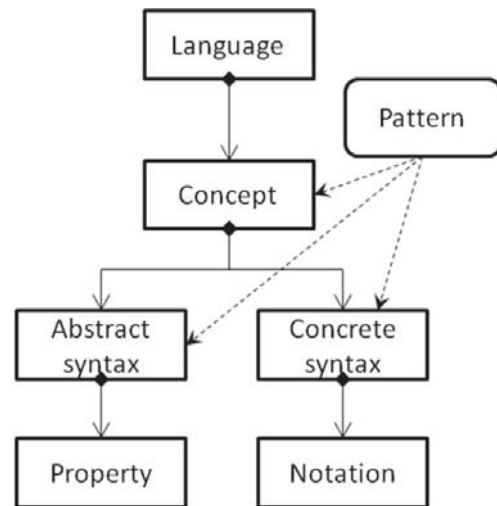
Fig. 1. Common language model

With the new model to internally represent designed language and provide checkable constraints and rules, we have achieved higher level of abstraction. Using this model it is easier to create more language specific tools, not just language processor as in our original modeling framework. In order to support previous version of describing language in annotated Java classes we have created tools for conversion between such language representation and representation in common language model. We will still be using Java classes as language implementation and as base for most tools.

V. LANGUAGE TOOLS

This section describes possible or existing tools for assisted language creation and maintenance. Some of tools can be created generically to support each language described using common language model and some tools can be generated specifically for designed language.

A. Input forms

One of valued property of designing new language is to be able to do so in more possible formats. This way you are not limited to usage of only one form, but you can choose from more or even create your own input form.

Input for common language model can be in form of language implementation written as Java classes for each concept and annotated with additional information for description of concrete syntax and parsing specific data. This option provides possibilities to use your existing implementation containing already semantic of each concept. Description of semantics of new language can be done inside language implementation even for other forms of language design input.

Second form of input is textual form for describing language concepts and syntax using our own domain specific language created for describing language's concepts, abstract and concrete syntax and assigning different language patterns to them. Partial example of description of language using our DSL is succeeding. Each pattern is written using curly brackets, concept starts with keyword *concept*, abstract syntax starts with keyword *AS* and concrete syntax starts with

keyword *CS*.

```
concept Constant
    AS: name : string {Identifier}, number : Number
    CS: name "=" number

concept Expression {Parentheses}

concept Factor : Expression

concept Number : Factor
    AS: value : int
    CS: value
```

Next option for defining language is possibility of using already existing language sentences and describing only names and abstract syntax of concepts used in language. Then we should be able to use syntax analyzer to analyze our sentences, find patterns in them and generate concrete syntax for language. Diagram of possible inputs is in Fig. 2. displayed as inputs to language model (upper part of figure).

In common language model we have created patterns, which can provide more information for item they are attached to. We can have patterns for language concepts, notation, notation parts, and properties of abstract syntax. Such patterns can describe constrains, references, formatting etc. For this kind of information in annotated classes we have corresponding annotations. Conversion between annotation and patterns is done automatically when creating language model.

B. Language implementation

In case you do not use annotated Java classes for design of new language it is expected to generate such classes automatically as they are needed for functional implementation of new language. We have designed and implemented generator, which creates one class for each specified language concept. For each possible language construct in language model there is defined its corresponding representation in annotated Java classes.

C. Language tools

Common internal language model can gain its value with possibilities of automatic creation of tools for each language. Such feature can even help in case of language evolution, when you change some language concepts and have to create completely new set of tools.

In the beginning we have identified some tools, which we are able to generate automatically. The most needed tool for textual languages is parser, which allows us to read and transform textual representation of language sentence to its memory representation. During this process we can try to find the best possible way of parser implementation using pattern recognition in required concrete syntax and determine if it would be better to use LL, LR or other type of parser.

In case that we have language allowing name references, we need to create tool for reference resolving. During reference resolve process we create directed graph from tree input. To make this part interoperable with different parsers, we have tried to implement it in aspect oriented solution, specifically AspectJ [12][13].

Traversing of created directed graph consisting of instances of language concepts can be traversed in simple manner. To provide language designer with such functionality we have created generator for tool which resembles Visitor pattern [9][10]. We can call this tool Concept visitor, as it goes
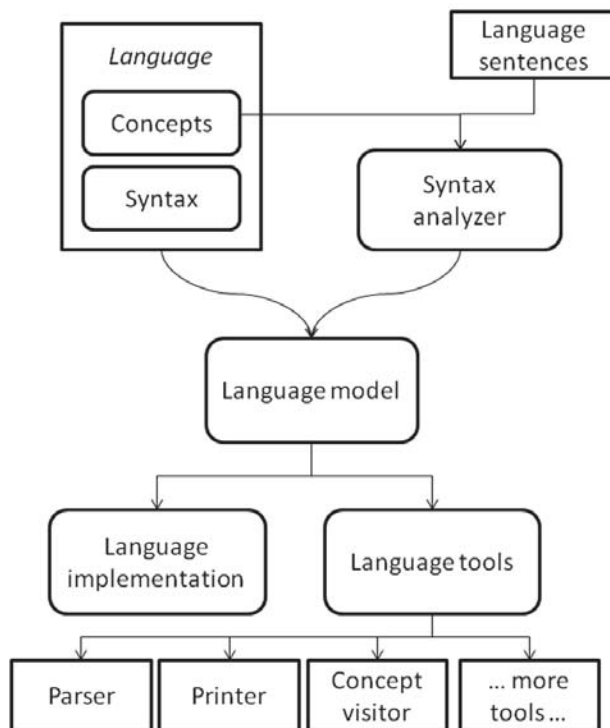
Fig. 2. Inputs and outputs of common language model

through each of concepts and its connected concepts.

Creation of textual representation from memory is inverse functionality of parser. It can be used to serialize information in memory. Thanks to concrete syntax part in our common language model, we can create tool for such serialization. This printer actually support pretty printing technique [11][14][15][16]. Generated printer is based on Visitor pattern. There were some problems needed to solve for creation of printer, which are out of scope of this article. For example we had to add patterns for defining formatting of pretty print output and solve decision for usage of best possible notation.

There can be even more tools created from common language model. For example using language syntax you can create plugin for syntax highlighting in IDEs.

VI. CONCLUSION

The aim of this paper was to present problematic of creation of new software languages and provide ideas for improving this process. We have also presented our conception for common language model. This paper presented design and ways of implementation of new tools for language oriented software engineering using our solution. Everything we described was also successfully implemented and tested on created experimental languages. Presented common language model and tools generated using this model, can save us from writing hundreds of source code lines and considerably decrease cost for creating and maintaining new languages.

REFERENCES

[1] G. L. Steele, *Growing a language*. In Addendum to the 1998
     Proceedings of the Conference on Object-Oriented Programming,
     Systems, Languages, and Applications (Addendum) (Vancouver,
     British Columbia, Canada). J. Haungs, Ed. OOPSLA '98 Addendum.
     ACM Press, New York, NY.

[2] F.P. Brooks, *Mythical Man-Month* (Anniversary Ed.). 1995. Addison-
     Wesley Longman Publishing Co., Inc.

[3] S. Freeman, N. Pryce, *Evolving an Embedded Domain-Specific
     Language in Java*. OOPSLA '06 October 22-26,2006, Portland,
     Oregon, USA.

[4] G. Gupta, *Language Based Software Engineering*. USA, Dallas:
     University of Texas. Department of computer science.

[5] R. Pressman, *Software Engineering: A Practitioner's Approach*. 4th
     Edition, McGraw-Hill, 1996.

[6] G. Kiczales, et al., *Aspect-Oriented Programming (AOP)*. ACM
     Computing Surveys, 28A, 4 (December). 1996.

[7] J. Porubän, *Návrh a implementácia počítačových jazykov*, Habilitation
     thesis. Košice: TUKE FEI KPI, 2008.

[8] J. Rumbaugh, I. Jacobson, G. Booch, *The Unified Modeling Language
     Reference Manual*. 2nd Edition. Addison-Wesley Professional, 2004.
     752 p.

[9] E. Gamma, et al., *Design Patterns*. Addison Wesley, 1995.

[10] R. C. Martin, et al., *Pattern Languages of Program Design 3*. Addison
     Wesley, 1998.

[11] G. Blaschek, J. Sametinger, *User-Adaptable Prettyprinting*, University
     of Linz, 1989.

[12] R. Laddad, *AspectJ in Action*, Manning, 2003.

[13] J.D. Gradecki, N. Lesiecki, *Mastering AspectJ*, Wiley, 2003.

[14] M. de Jonge, *Pretty-Printing for Software Reengineering*, icsm,
     pp.0550, 18th IEEE International Conference on Software Maintenance
     (ICSM'02), 2002

[15] D. Oppen. *Prettyprinting*. ACM Transactions on Programming
     Languages and Systems, 2(4):465–483, 1980.

[16] M. de Jonge. *A pretty-printer for every occasion*. In I. Ferguson, J.
     Gray, and L. Scott, editors, Proceedings of the 2nd International
     Symposium on Constructing Software Engineering Tools
     (CoSET2000). University of Wollongong, Australia, 2000.

BASIS FOR GUI DOMAIN USABILITY ANALYSIS: FORMALIZATION

[1]*Michaela KREUTZOVÁ,*

[1]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]michaela.kreutzova@tuke.sk

Abstract—**Domain analysis is an effective technique for enabling many features including reuse, reverse engineering and usability analysis. This refers to GUIs (Graphical User Interfaces) too, which is the most important part of software allowing user interaction. This paper describes how GUI formalization can provide a basis for automatic domain analysis and later for usability analysis. It describes different types of components and their meaning as formalized domain terms, their properties and relationships between them. The domain terms are then recorded automatically from the GUI into a GUIIL domain specific language. This paper also defines two hypothesis for making domain analysis possible, determines the conditions, under which these goals can be fulfilled.**

Keywords—**usability, domain-specific language, graphical user interface, component-based analysis, Java.**

I. INTRODUCTION

Domain analysis - the process of collecting and organizing information about a specific class of problems - has primarily been a technique for facilitating software reuse [1]. The structure of a defined domain can provide the basis for retrieving information about software components, for managing complexity and for making reuse more efficient and effective [2].

A. Motivation

When evaluating application usability, most of the tests and evaluations focus on application fulfilling requirements of *usability definition*, for instance like in the approaches described in [3], [4].

When evaluating graphical interface usability, we concentrate on two main aspects affecting usability:

1) *Ergonomic* - main factors are: component positions in the interface, colors, size of components and fonts, etc.
2) *Content-procedural*

The (1) refers to the terms that can be found in the interface. They can be *text-based* or *image-based*. For the user, most important is to *understand* them. Second important thing is, that the terms contained in the interface either should be common, or they should belong to the application domain. Let's imagine a pharmacist or doctor working with an application attempting to prescribe a receipt for his patient allowing him to buy medicine in a pharmacy. In order to be able to work with the application, he should see terms like "patient", his "name", "surname", "birth number" and terms like "medicine", its "bar code", "name", "price", "usage", "volume" or "number of pills", "ingredients", etc. All these terms belong to the domain of medicine and pharmacy. But the doctor should not
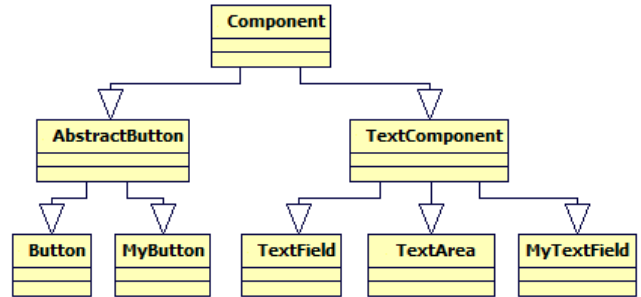
Fig. 1. Hierarchy of the components is created by inheritance

see terms like "My bank account", "Author of a book" or "Building block" in his application. These terms don't belong to a domain of medicine and pharmacy.

We think that these two parts - understandability and relation to the domain are very important when evaluating usability.

B. Problems & Objectives

At present there are some procedures for GUI usability evaluation are focusing on interface *understandability* and on the terms domain-relation. We researched mainly in the area of GUI usability evaluation (approaches described in [3], [4]) and, because it is very popular and it is not very far from usability evaluation, testing ([5], [6], [7], [8], [9], [10], [11], [12], [13], [14]).

Out of the researched approaches, only one evaluates spelling of the terms (as it is with Sherlock described in [3]) or the terms are not analyzed at all. But none of the researched approaches analyzes the correct *semantics* of the terms.

Our goal is to propose an approach for evaluating application usability in terms of domain concepts and a system that (in it's final state) would be able to assess (to some extent) understandability of terms in application - their classification to the application domain, relationships between them and their connection to their organization on components and layout in the application (hierarchy of terms - see fig. 1).

Our final goal is to make automatic usability evaluation possible. But first we have to prove it's even possible. Therefore we state two hypothesis:

1) The graphical user interface is a (partial) description of the application domain.
2) It's possible to (automatically) analyze existing user interface and search for domain terms in the interface.

Based on the hypothesis we state the following goals:

1) Prove that hypothesis 1 is valid.

2) Based on hypothesis 1 prove that hypothesis 2 is valid.
3) Create a software solution to realize an experiment of domain terms formalization as basis of application user interface usability evaluation

We mentioned the first assumption in 1.1. For instance, a bank system uses terms from banking. So we can say, that for the application, in order to be *understandable* for users, hypothesis 1 *must* be fulfilled.

II. RELATED WORK

In Sofia a research in information understandability in interface design [15] is made, which is similar to our work. But it's on a more ergonomic level. Earlier work [2] also deals with user interface formalization and its subsequent transformation to a different interface. But they're doing the whole procedure on a console interface, which is considerably easier to formalize than a graphical user interface.

III. BACKGROUND

A. Methodology

It's possible to use one of these three basic types of usability analysis:

- *Component-based*
- *Image-based*
- *Source code based*

The methodology of aplication domain terms analysis is depending on the application structure: if the application is made of components, we use component-based analysis; if it's not possible, we must analyze the interface from the graphical aspect (graphics recognition, images, fonts, etc.) like it is described in [16]; and if we can not use use either first or second approach we go to the source code (source code analysis), i.e. like in [2], [6] and [17]. For our approach, in order to use the image-based analysis, we would need a graphics recognition software to recognize the terms in the interface and it would be very complicated or impossible to recognize non-standard components and their relation to the terms. And since we're trying to analyze the *user interface*, we can not go into the source code either. Therefore we decided to use the first approach to prove the second hypothesis. During our work several questions raised:

- Is the component-based analysis possible?
- When? What are the restrictions?
- How to identify components?
- How to extract data out of the components?

The component-based analysis is possible only if the application is *made of components*. As for the other three questions, we will answer them in the following sections.

B. Domain terms in GUI

We mentioned that in order to use component-based approach the GUI should be made of components. A basis is always some window or a dialog and we put every component in it. For better prospect we can put the components into unlimited amount of containers (`Panel`). All the components and containers in the interface represent *terms* in the user interface. For these terms there exist some specific *rules*, there are specific *relationships* between them and their *hierarchy* is exactly determined. For instance, `Button` and `Panel`
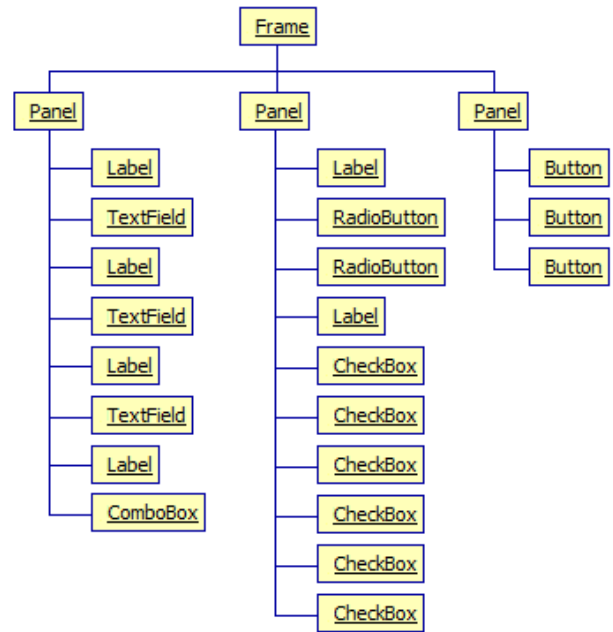
Fig. 2. Person form and mapping its components into a component tree

represent terms. Their relationship is that a `Panel` can contain `Button` components. A hierarchy builds up using inheritance, which defines a membership of the component in a logically related group. For instance if `JButton` inherits from `AbstractButton`, this relationship automatically puts it into a group of buttons, assigning correct attributes and functions to it and determining a correct interface for its class.

We can identify these *terms*, extract information from them and reflect them into a *"component tree"* (Figure 2). Name of the window/dialog can tell us, to which domain (domain part) its components belong (for instance on Figure 2 the components represent information about a *Person*).

C. Domain-Specific Unit

The domain specific units are components. The following list presents the most frequently used graphical components of Java Swing library in alphabetical order and their meaning as *terms* of GUI language:

- *AbstractButton* - abstract class representing all functional components of button and menu type. Buttons and menus are basic functional components of a GUI and they relate to a window (dialog), in which they are located. List of all buttons and menus in a window (dialog) represents *basic functionality* of the window (dialog). Excluding general terms like OK, Cancel, Edit, Add, Delete, Save etc.) we are getting *domain-specific application functionality*.
- Container - represents a grouping of components. About domain-specific components located in a container we can tell that they *are related logically or graphically*.
- ComboBox - allows choosing one item of a list. Items can represent domain-specific units and we can tell that they *are logically related* and *mutually exclusive*.
- FileChooser - allows *choosing a file or directory*. It's a functional GUI item enabling using a local file or directory to read or process.
- List - allows choosing one or more items of a list. Items can represent domain-specific units and we can tell that

they are *logically related* and *not mutually exclusive* - it's possible to choose more items at once depending on selection model.

- Menu - a specific case of a container grouping together menu items (or functional elements). We can tell that items of its group are *logically related in functionality*. For instance a menu File groups together functional elements allowing file manipulation like: Open, Save, Save As..., Close etc. A menu can represent a group of logically related functional domain-specific units.

- Spinner - represents a functional item displaying a specific type of data (number, date etc.). It allows incrementation and decrementation of these data using its functional buttons (arrows). About domain-specific units of this type we can tell what type they are (according to display) and what their *constraints* are.

- TabbedPane - a specific type of container allowing displaying its components on one or more switchable pages (tabs). These pages group together *logically related* components. Thus a JTabbedPane component can represent a list of groups of domain specific units and membership of a unit is determined with a tab title.

- TextComponent - represents all basic text components. About this item type we can tell that it *serves for acquiring and/or saving data of text* type given by user. We can get information about *type of data* (from label, type - classic text field, password text field etc.), their *maximal length*, about their *content* and *constraints* for it; and in case of existence of a label also about their *purpose*.

- Tree - represents a tree structure. The tree label tells us, *to which domain do its items belong*. Items of the tree represent *units* of this domain (domain part - subdomain) with a precise *hierarchical order*, which indicates *relations* between them.

- Own components - when creating new components a programmer *inserts new terms into the general language*, he enlarges the general language with new domain-specific terms. A domain-specific content depends on a concrete component implementation.

[18] says that a programmer creates a new component, it is like an injection of new words into natural interface vocabulary. However, if a programmer creates these components incorrectly (he defines terms, relationships and rules incorrectly or not at all) it's like a violation of a natural language, inserting unknown terms in it making impossible further communication in this language.

The components therefore should be created according to a component definition so that they'd be fully open, reusable [19], [20] and so they would not interfere with the existing GUI language.

D. Identification of domain specific units

Different types of components are identifiable differently and we can extract different information from them. Components in the component tree can be identifiable with these basic properties:

1) *Component type* - a class of the component.
2) *Component name* - the name attribute of the component.
3) *Component title* - the title attribute of the component.
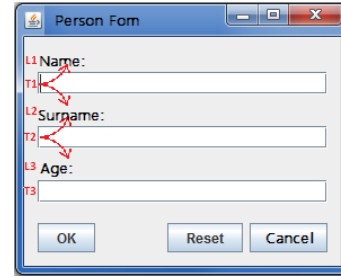
Fig. 3. Person form and labelFor attribute problem

4) *Component tool tip text* - the toolTipText attribute of the component.
5) *Action command* - the actionCommand attribute of a functional component.
6) *A labelFor* attribute of a Label component - when using a component, that doesn't have its own visible identifier we can use a Label and set its labelFor attribute to refer to that component.
7) *Component path* - a path in the component tree. It is a representation in memory and that always exists. Therefore a computer path is a universal identifier.
8) *Component coordinates* - not stable, depends on interface shape, component layouts and many other conditions.
9) *Own components* - specific component identification capability depends on correct component implementation.

Points 1 to 6, if a component is identifiable by one of them, represent domain-specific information. If a component is not identified by any of the attributes 1 to 6, we determine it as a potential usability issue. In this case, we use component path and coordinates to identify this component.

Based on terms presented in section 6 and identification properties presented in this section an application domain analysis can be conducted. Figure 2 shows problem with labels and components. Without the labelFor attribute filled, we can't say if the T1 text field belongs to L1 or L2 label. So it is with T2.

IV. FORMALIZATION SOFTWARE

A software solution realizing formalization of GUI domain terms was implemented. This software solution uses a component-based approach and the principle described in the previous chapters to identify (and work with) components representing terms of the interface. It proves the hypothesis 2 - the terms in the interface can be formalized. Conditions for this approach are described in 4. The components are identified using their textual identificators visible in the interface. This is crucial, because even if the whole interface would be replaced by another, the terms would be still there and our software would be still usable. Input and output files for our software solution are recorded in a simple domain-specific language GUIIL. A DSL is a programming language of limited expressiveness focused on a particular domain (for instance medicine, banking, building, etc.) [21], [22], [23]. A sentence in this language consists of a command sequence that can be executed on GUI [24].

V. EXPERIMENTS & RESULTS

Experiments were made on few open-source applications - *jEdit*, *GuitarScaleAssistant*, and *SweetHome3D*.

The results showed that there are some obstacles to our approach:

1) The application has to have common components, best fulfilling the Java look & feel guidelines [25] in order to be able to identify components using our approach. One of related problem is the labelFor attribute issue described on fig. 3.

2) If the application has its own class loader, it's impossible to identify components using our software solution.

3) Every component has to have its description or label to be able to identify it as a domain term.

The solution for both 1) and 2) would be to create handlers using aspect oriented programming, either for the custom components or for the components that are not accessible because of the second class loader. Point 1) is considered as a potential usability issue. Point 3) describes the basic mistake made by programmers, not identifying the components and this is a basic assumption for understandability. Therefore we also consider it to be a potential usability issue.

VI. Conclusion

We presented here a basis for a domain driven graphical user interface analysis: a formalization of terms situated in the interface. We defined a domain-specific language for user interface and thus we enable component identification based on domain terms that they represent. We also defined conditions for such identification. Problems arised in process of identifying components were either caused by incorrect implementation of components (this we stated as a potential usability issue) or by the application's own class loader. As a solution we proposed a way to create new drivers using aspect oriented programming.

In the future we plan further research in the area of automatic application usability evaluation. Possible challenges are also methodical learning process of software usage and automatic generation of software manuals.

Acknowledgment

References

[1] R. Prieto-Diaz and G. Arango, *Domain Analysis and Software Systems Modeling*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1991.

[2] M. M. Moore and S. Rugaber, "Domain analysis for transformational reuse," in *Proceedings of the Fourth Working Conference on Reverse Engineering (WCRE '97)*. Washington, DC, USA: IEEE Computer Society, 1997, pp. 156–. [Online]. Available: http://portal.acm.org/citation.cfm?id=832304.837003

[3] M. Y. Ivory and M. A. Hearst, "The state of the art in automating usability evaluation of user interfaces," *ACM Comput. Surv.*, vol. 33, pp. 470–516, December 2001. [Online]. Available: http://doi.acm.org/10.1145/503112.503114

[4] D. M. Hilbert and D. F. Redmiles, "Extracting usability information from user interface events," *ACM Comput. Surv.*, vol. 32, pp. 384–421, December 2000. [Online]. Available: http://doi.acm.org/10.1145/371578.371593

[5] X. Yuan and A. M. Memon, "Iterative execution-feedback model-directed GUI testing," *Inf. Softw. Technol.*, vol. 52, pp. 559–575, May 2010. [Online]. Available: http://dx.doi.org/10.1016/j.infsof.2009.11.009

[6] R. Lämmel *et al.*, "Large-scale, AST-based API-usage analysis of open-source Java projects," in *SAC'11 - ACM 2011 SYMPOSIUM ON APPLIED COMPUTING, Technical Track on "Programming Languages"*, 2011.

[7] X. Yuan and A. M. Memon, "Generating Event Sequence-Based Test Cases Using GUI Runtime State Feedback," *IEEE Trans. Softw. Eng.*, vol. 36, pp. 81–95, January 2010. [Online]. Available: http://dx.doi.org/10.1109/TSE.2009.68

[8] M. Umar, "Generating Automated Test Scripts for AltioLive using QF-Test IntegraSP Ltd." Integra SP. London, United Kingdom, 2008.

[9] P. A. Brooks and A. M. Memon, "Automated gui testing guided by usage profiles," in *Proceedings of the twenty-second IEEE/ACM international conference on Automated software engineering*, ser. ASE '07. New York, NY, USA: ACM, 2007, pp. 333–342. [Online]. Available: http://doi.acm.org/10.1145/1321631.1321681

[10] W. M. Li, Kanglin, *Effective GUI Test Automation: Developing an Automated GUI Testing Tool*. SYBEX Inc., 2004.

[11] M. Last *et al.*, *Artificial Intelligence Methods In Software Testing (Series in Machine Perception & Artifical Intelligence " Vol. 56)*. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2004.

[12] A. M. Memon *et al.*, "Hierarchical GUI Test Case Generation Using Automated Planning," *IEEE Trans. Softw. Eng.*, vol. 27, pp. 144–155, February 2001. [Online]. Available: http://portal.acm.org/citation.cfm?id=373394.373399

[13] S. McMaster and A. Memon, "Call-Stack Coverage for GUI Test Suite Reduction," *IEEE Trans. Softw. Eng.*, vol. 34, pp. 99–115, January 2008. [Online]. Available: http://portal.acm.org/citation.cfm?id=1340674.1340716

[14] A. M. Memon *et al.*, "Plan Generation for GUI Testing," in *Proceedings of The Fifth International Conference on Artificial Intelligence Planning and Scheduling*. AAAI Press, Apr. 2000, pp. 226–235.

[15] S. Vassileva and S. Anguelova, "Research on information understandability in interface design," 2003, http://www.create.uwe.ac.uk/membership/VassilevaAnguelovaInformationUndestandi

[16] J. Lladós and Y.-B. Kwon, "Graphics Recognition: Recent Advances and Perspectives," *Lecture Notes in Computer Science*, vol. 3088, pp. 291–363, 2003.

[17] M. Acharya *et al.*, "Mining API patterns as partial orders from source code: from usage scenarios to specifications," in *Proceedings of the the 6th joint meeting of the European soft. eng. confer. and the ACM SIGSOFT symposium on The foundations of soft. eng.*, ser. ESEC-FSE '07. New York, NY, USA: ACM, 2007, pp. 25–34. [Online]. Available: http://doi.acm.org/10.1145/1287624.1287630

[18] J. Nielsen, "Top-10 Application-Design Mistakes," http://www.useit.com/alertbox/application-mistakes.html, 2008.

[19] M. Slašťan and P. Mederly, "Software Reuse," Lecture for subject Software engineering, University of Komenský in Bratislava, http://cyril.fmph.uniba.sk/www/sw-eng/03-04-zs/, 2003.

[20] C. W. Krueger, "Software reuse," *ACM Comput. Surv.*, vol. 24, pp. 131–183, June 1992. [Online]. Available: http://doi.acm.org/10.1145/130844.130856

[21] Fowler, Martin, *Domain Specific Languages*. Addison-Wesley Professional (ISBN 0321712943), 2010.

[22] M. Mernik *et al.*, "When and how to develop domain-specific languages," *ACM Comput. Surv.*, vol. 37, pp. 316–344, December 2005. [Online]. Available: http://doi.acm.org/10.1145/1118890.1118892

[23] P. Hudak, "Modular Domain Specific Languages and Tools," in *Proceedings of the 5th International Conference on Software Reuse*, ser. ICSR '98. Washington, DC, USA: IEEE Computer Society, 1998, pp. 134–. [Online]. Available: http://portal.acm.org/citation.cfm?id=551789.853532

[24] J. Porubän, "Design and implementation of computer languages," Master's thesis, Technical University of Košice, 2008.

[25] Sun Microsystems, Inc., "Java Look and Feel Design Guidelines," 1st edition, Javasoft, 1999.

Brief overview about broadcast news audio segmentation, classification and retrieval system

Jozef VAVREK

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

jozef.vavrek@tuke.sk

Abstract—**Audio segmentation and classification is an essential part of broadcast news processing system. This paper describes basic concept of such content-based analysis and retrieval system. The basic idea of designing an effective audio stream processing system is to distinguish four fundamental sound classes, namely speech, music, environment sound and silence. There exist several techniques and algorithms that ensure correct segmentation and classification of each single type of class, but discrimination of mixed type audio data is still considered a difficult problem. The effort is therefore aimed to develop fast, efficient and reliable classification system with ability of adaptive multi-class classification for various types of speakers and acoustic events. One of the most substantial step in creating such classification system is selection of an optimal feature set and robust classifier. These two factors significantly influence resulting classification accuracy. Retrieval accuracy of the proposed system depends beside mentioned methods on indexing or annotation of audio content as well.**

Keywords—**Content-based analysis, broadcast news audio stream, feature selection, segmentation, classification, indexing and retrieving.**

I. INTRODUCTION

Rapid increase in the amount of audio data demands for efficient and fast processing methods that allow automated classification and retrieval of broadcast news database. All these content-based techniques have been used in order to ensure audio content processing in terms of discrimination various types of speakers and acoustic events as well. A fundamental step for processing audio stream is to automatically classify or divide an audio content into homogeneous segments. We call this separation criterion as segmentation. Segmentation is substantial part of whole proposed system and is very useful in many classification task. Segmentation is usually carried out along with the process of future extraction.

This paper describes basic concept of classification and retrieval system used for processing broadcast news data, a mass media with rich audio content. Audio content of broadcast news media contains not only single type of classes (pure speech, music), but also mixed sounds (speech and music with background). That audio stream consists of the following audio segments:

- *pure speech* - single anchor reports in studio, field speech in a clean environment, telecaster, telephone conversations between two subjects,
- *speech with environment sound* - field speech or interviews in noisy environment,
- *speech with music background* - program titles and program closings, commercials,
- *pure music* - program titles, music in commercials,
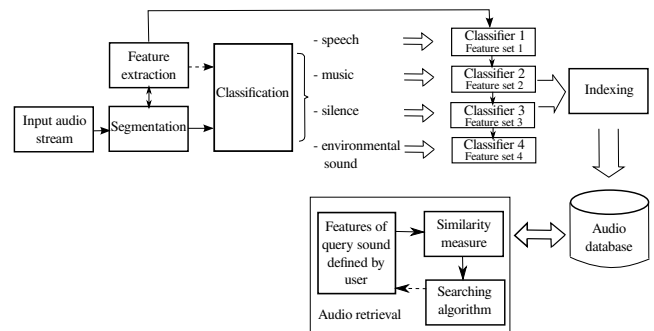
Fig. 1. Basic concept of automated classification and retrieval system for broadcast news audio stream

- *environment sound* - all types of acoustic events with field occurance, e.g. sound from machines, birds, water, wind, crowds, etc.,
- *silence* - pause intervals between different speakers and news stories.

The main effort in content-based audio analysis systems tends to separation two types of events, namely speech and music (non-speech). A human listener can easily recognise various types of speech and music by listening only small parts of audio signals, but it's very difficult problem for computer programs.

Zhang and Kuo used hidden Markov models (HMM) to distinguish speech, music and environmental sound with various types of feature extraction methods [1]. The main idea of their proposed system consists in rule-based heuristic procedure which means, that whenever there is an abrupt change detected in any of the basic features, a segment boundary is set. In addition, Han, Gao and Ji focused on the application of Support Vector Machines (SVM) method for classification and detection audio signal by means of new proposed method, namely selective ensemble SVM with much more used features [2]. Other works are aimed on the design of such complex systems, that are able to process broadcast news audio signals, in terms of segmentation, classification, indexing and retrieval [3], [4], [5]. Some authors have developed new and efficient features and segmentation algorithms that can capture all possible changes in audio stream, thus improve classification accuracy of audio data [6], [7]. The basic scheme of the most common used classification and retrieval systems is illustrated in Fig. 1.

This paper is organized as follows: In section II are introduced audio features and segmentation methods, section III provides an overview about used classifiers and section

IV describes some characteristics about indexing and retrieval methods. Finally, the section V gives our conclusions and future work proposal.

II. Segmentation and Feature Selection

The primary and important task in broadcast news segmentation, classification and retrieval is to extract features that objectively represent the audio content. All the possible and often used features can be categorized into four classes [8]:

- *Energy features* - short time energy (STE), silence frame ratio (SFR), noise frame ratio (NFR), subband energy distribution (SED), percentage of low-energy frames and modulation energy,
- *Statistical spectral features* - more of them are defined in standard MPEG-7 [9], for example spectral roll-of of the centroid,spectral flux, spectral kurtosis, zero-crossing rate and ratio, audio spectrum spread and flatness, etc.,
- *Cepstral features* - are derived from human perceptual (auditory) system, such as Mel Frequency cepstral co-efficients (MFCC), linear predictive cepstral coefficients (LPCC), linear spectral pairs (LSP), perceptual linear predictive analysis (PLP) and mean of minimum cepstral distance (MMCD),
- *Pitch features* - such as pitch and delta pitch, spectral peak duration, pitch tuning, average pitch-density (APD) and relative tonal power density (RTPD).

Feature extraction is substantial part of audio-content processing, but we must realize segmentation of audio stream too. Audio segmentation in general is the task of segmenting a continuous audio stream into acoustically homogeneous regions, where the rule of homogeneity depends on the discrimination criterion. In other words, segments are short audio portions, usually with equal length (1 - 5s), that contain the same type of acoustic signal (speech, music or environmental sound). Many of works and studies are focused on development an alternative segmentation methods. For example, Huang and Hansen have proposed a novel segmentation and classification algorithm *CompSeg*, that can achieve effective unsupervised speaker segmentation, especially for short duration segments [10]. It combines three distance metrics techniques, namely Bayesian information criterion (BIC), T^2 distance and weighted mean distance (WMD). Siegler at al. used the Symmetric Kullback-Leibler distance for speaker segmentation [11], but this method is effective only in a case, when length of segments is longer than 5s. Lin at al. have developed an STMR algorithm for identification dissimilarity boundaries between different speakers, which uses the principles of unsupervised segmentation by SVM [12].

All mentioned works and many others are based on the principles of supervised and unsupervised audio content change detection (ACD) [13]. Therefore segmentation ability strongly depends on the segment length of the audio stream and assumption, that the acoustic data are composed of different speakers who are either known a priory or unknown.

III. Classification

Audio classification has been an active research area in recent years. The performance of an audio classification depends primarily on the effectiveness of the segmentation and used feature extraction methods. A basic idea of classification broadcast audio stream lies in two steps. At first speech,

music, environment sound and silence are separated. Then more specific classes and unknown sounds are distinguish within each basic audio type. Thus we try to build an optimal and efficient training model for each separated sound class. Another step within solving a classification issue is the selection of appropriate classifier. The most popular and often applied classifiers are hidden Markov model (HMM), Gaussian mixture model (GMM), multi-layer perceptron (MLP) and other neural networks (NN), K-nearest neighbour (KNN), Gaussian likelihood ratio (GLR), Bayesian networks (BN) and support vector machines (SVM). Liu, Xie and Meng have experimented with three types of classifiers KNN, MLP and SVM and developed an audio classifier that discriminates between speech and music segments in Mandarin broadcast news [14]. Dhanalakshmi at al. have proposed an automatic audio classification system using new auto associative neural network algorithm, based on feed forward neural networks [15]. Other works use combination of several types of classification techniques with various types of feature extraction methods [16], [17], [18].

SVMs [19] have in recent years drawn much interest for audio classification due to their generalisation ability and superior performance in various pattern classification tasks. Many of works and studies have discovered that SVMs outperform other types of classifiers in terms of classification accuracy. For example, in Microsoft research [20], SVM was used for audio stream classification into one of the five classes: pure speech, non-pure speech, music, environment sound, and silence. Experimental evaluations have showed that the performance of SVM is much higher than that of using KNN and GMM whatever the testing unit length was. But the use of SVM for multi-class audio classification isn't simple task. This is because the SVM is originally designed for solving binary classification problem and discrimination of multiple classes is realized by combining several binary classifiers. Therefore we try to create an effective classification binary tree scheme, in order to increase total performance of SVM technique [21].

IV. Indexing and Retrieving

The last, but essential step in the process of content-based audio classification and retrieval is *indexing*, or *annotation* and subsequent *retrieval* of broadcast news audio stream. The main task of indexing methods is to provide labeling of audio stream with specific labels in terms of audio content transcription. Sounds within the individual classes are then retrieved by these characteristic labels and ordered according to similarity with the query sound. We call this retrieval system as *query-by-example* audio retrieval [1], [22]. Query-by-example methods are based on similarity measures between audio example given by the user and sounds that are located in a database. In audio retrieval, most conventional methods are based on Euclidean or Mahalanobis distance to classify the query sounds into one of the sound in the database.

A promising method for retrieving query examples in broadcast news database is called distance-from-boundary (DFB) technique [23]. It uses an SVM approach when the system firs finds a boundary inside which the query pattern is located and then are all the audio patterns in the audio set sorted according to their distances to this boundary.

V. Conclusion and Future Work

A brief overview about content-based broadcast news segmentation, classification, indexing and retrieval was presented in this paper. There is some description about feature extraction methods and segmentation algorithms used to process audio streams from TV news in current technology development. This work also captures basic ideas of classification methods and their use for the final retrieving in audio databases.

Future work will be led to build an effective and robust classification and retrieval system for broadcast news audio stream. At firs, we will solve a problem of selection an appropriate feature extraction method. There exist several variables and feature selection methods, which select an optimal feature subset according to their usefulness to the classifiers, for example variable ranking, wrappers, filters, embedded, and nested subset methods [24]. We will choose one of these selective methods and use an effective segmentation algorithm to divide each audio stream into non-overlaping segments, for example segmentation by sliding window. Our work will continues, after segmentation and feature extraction, with selecting an optimal classifier. We will use the SVM method as a classifier, because of high efficiency, robustness and previous work with this technique [25]. Chosen classification method will provide discrimination of all possible sounds in audio stream and should give the condition for further processing at indexing and query-by-example retrieval stage. All future experiments will be realized by broadcast news database KEMT-BN, which contain over 100 hours recorded audio stream from TV news programs. We will use the open-source *Marsyas*[1] toolkit for feature extraction, SVM training and classification experiments.

The main effort will tend to propose an effective broadcast news audio classification and retrieval system with ability of adaptation to various types of acoustic events and speakers. For this purpose we want to develop a new segmentation and classification scheme with optimal feature set for each binary classifier. Our work will be also aimed to find efficient indexing and retrieval methods for increasing overall performance of proposed system.

Acknowledgment

References

[1] T. Zhang and C. C. Kuo, "Hierarchical classification of audio data for archiving and retrieving," *ICASSP '99 Proceedings of the Acoustic, Speech, and Signal Processing*, vol. 06, pp. 3001–3004, 1999.

[2] B. Hang, X. Gao, and H. Ji, *Advances in Neural Networks - ISNN 2005*. Springer Berlin / Heidelberg, 2005, vol. 3497, ch. Automatic News Audio Classification Based on Selective Ensemble SVMs, p. 811. [Online]. Available: http://dx.doi.org/10.1007/11427445_59

[3] T. Zhang and C.-C. Jay Kuo, "Audio content analysis for online audiovisual data segmentation and classification," in *IEEE Transactions on Speech and Audio Processing*, May 2001, vol. 9, pp. 441–457.

[4] R. van Dinther, M. F. McKinney, and H. R. Li, "Real-time segmentation of rad io broadcast content in radio devices," *Computers in Education, International Conference on*, vol. 0, pp. 1–2, 2009.

[5] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification," in *in Proc. SIGIR, 2003*, 2003, pp. 282–289.

[6] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," in *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)-Volume 2 - Volume 2*, ser. ICASSP '97. Washington, DC, USA: IEEE Computer Society, 1997, pp. 1331–. [Online]. Available: http://portal.acm.org/citation.cfm?id=839274.839391

[7] M. Ramona and G. Gichard, "Comparison of different strategies for a svm-based audio segmentation," in *17th European Signal Processing Conference (EUSIPCO 2009)*, ser. EUSIPCO 09, 2009. [Online]. Available: http://quaero.org

[8] L. Xie, Z.-H. Fu, W. Feng, and Y. Luo, "Pitch-density-based features and an svm binary tree approach for multi-class audio classification in broadcast news," *Multimedia Systems*, vol. 17, pp. 101–112, 2011, 10.1007/s00530-010-0205-x. [Online]. Available: http://dx.doi.org/10.1007/s00530-010-0205-x

[9] M. Casey, "General sound classification and similarity in mpeg-7," *Organised Sound*, vol. 06, pp. 153–164, 2001.

[10] R. Huang and J. H. L. Hansen, "Advances in unsupervised audio classification and segmentation for the broadcast news and ngsw corpora," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 907–919, May 2010.

[11] M. A. Siegler, U. Jain, B. Raj, and R. M. Stern, "Automatic segmentation, classification and clustering of broadcast news audio," in *Proc. DARPA Speech Recognition Workshop*, 1997, pp. 97–99.

[12] P. C. Lin, J. C. Wang, J. F. Wang, and H. C. Sung, "Unsupervised speaker change detection using svm training misclassification rate," *IEEE Transactions on Computers*, vol. 56, pp. 1234–1244, 2007.

[13] C.-C. Huang, J.-F. Wang, and D.-J. Wu, "Automatic scene change detection for composed speech and music sound under low snr noisy environment," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 13, pp. 689–699, 2005.

[14] C. Liu, L. Xie, and H. Meng, "Classification of music and speech in mandarin news broadcasts," *9th National Conference on Man-Machine Speech Communication*, vol. 9, 2007.

[15] P. Dhanalakshmi, S. Palanivel, and V. Ramalingam, "Automatic segmentation and classification of audio broadcast data," *Asian Journal of Information Technology*, vol. 09, pp. 54–61, 2010.

[16] L. Chen, S. Guduz, and M. T. Ozsu, "Mixed type audio classification with support vector machines," in *ICME '06*, 2006, pp. 781–784.

[17] Y. Y. Chung, E. H. C. Choi, L. Liu, M. A. M. Shukran, D. Y. Shi, and F. Chen, "A new hybrid audio classification algorithm based on svm weight factor and euclidean distance," in *Proceedings of the 2007 annual Conference on International Conference on Computer Engineering and Applications*. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2007, pp. 152–157. [Online]. Available: http://portal.acm.org/citation.cfm?id=1348258.1348286

[18] P. Dhanalakshmi, S. Palanivel, and V. Ramalingam, "Classification of audio signals using svm and rbfnn," *Expert Syst. Appl.*, vol. 36, pp. 6069–6075, April 2009. [Online]. Available: http://portal.acm.org/citation.cfm?id=1497653.1498268

[19] S. Abe, *Support Vector Machines for Pattern Classification (Advances in Pattern Recognition)*. Secaucus, NJ, USA: Springer-Verlang New York, Inc., 2005.

[20] L. Lu, H.-J. Zhang, and S. Z. Li, "Content-based audio classification and segmentation by using support vector machines," *Multimedia Systems*, vol. 8, pp. 482–492, 2003, 10.1007/s00530-002-0065-0. [Online]. Available: http://dx.doi.org/10.1007/s00530-002-0065-0

[21] B. Fei and J. Liu, "Unsupervised speaker change detection using svm training misclassification rate," *IEEE Transactions on Neural Networks*, vol. 17, pp. 696 – 704, May 2006.

[22] M. Heln and T. Virtanen, "Audio query by example using similarity measures between probability density functions of features," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2010, p. 12, 2010.

[23] S. Guodong Guo; Li, "Content-based audio classification and retrieval by support vector machines," *IEEE Transactions on Neural Networks*, vol. 14, pp. 209–215, Jan 2003. [Online]. Available: http://dx.doi.org/10.1007/s00530-002-0065-0

[24] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[25] J. Vavrek, M. Pleva, and J. Juhar, "Acoustic events detection with support vector machines," in *Electrical Engineering and Informatics : Proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Koice*, 2010, vol. 1, pp. 796–801.

[1] http://marsyas.sness.net/

Building a Semantic map of science vocabulary using ontology mapping

Ján KAŽIMÍR

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

jan.kazimir@tuke.sk

Abstract—**In this paper I will describe steps, needs and reasons for developing an automatic system that will be able to create an semantic map of science vocabulary. Also I will subscribe a new path to go, in hope to challenge disadvantages with building a semantic web.**

Keywords—**Ontology mapping, semantic mapping, web agent.**

I. INTRODUCTION

In these times, there is too much knowledge stored on websites. For us, young researchers, is getting harder to tell what else from that we already know is relevant for our work. Instead of searching for known unknown we can start to search for the way to create an agent, which will be able to find this information for us. The biggest question is how that agent should work.

The answer is in our heads. One of interesting ways how our brain discovers new knowledge is pattern comparison of known and new things, situations, etc. Even slight match of patterns can lead to major discovery. And we will use this property.

Before we will be able to create such an agent, we must answer the question of how to find, compare and match informational patterns. It is also important, that we provide our new agent with capability of relative relevancy determination.

II. AGENT REQUIREMENTS

As I previously mentioned, the agent must match some requirements. Before I name and describe them, there is important to tell, that the agent we are talking about will by an extended Semantic web agent. Semantic web agents are able to retrieve some information, but it is still not enough. New, intelligent and one purpose agent should be able to:

A. Retrieve information

New agent will be still a Semantic web agent. That involves all its functionality. It includes specially marked information retrieval, multiple solutions and multiple sources retrieval. Those are usual requirements for an agent. New one should retrieve relevant, even not marked information. It should still retrieve multiple solutions from multiple sources.

B. Create ontology

The biggest challenge and base feature for new agent is to provide him with functionality of **Semi-automatic ontology creation**. This will be provided by altering of technique named Semantic mapping.

Semantic mapping is a visual strategy for vocabulary expansion and extension of knowledge by displaying in categories words related to one another. We will take the semantic mapping as a process of creation ontology of word terms and its graphical representation.

The process of ontology creation will be semi-automatic. It means that there is no system or even a concept of system that can provide automatic relevancy determination. So, because of this problem, we need at least one person, an expert, is needed to determine relevancy of partial results of process.

C. Represent knowledge stored in ontology

The last job for our agent is the representation and retrieval of learned knowledge. In our case it means, that we need to present only partial, the relevant part of ontology that was previously created. The relevant part will be determined by user query.

III. THE PROCESS

Now that we know the needs for an agent, we will describe the steps of the process from search for information to user query answer display. Whole process can be divided to two separated processes. First of them is knowledge base creation and second is knowledge retrieval.

Creating knowledge base

Process of Knowledge base creation is described as off-line process. It mean that it is independent and do not depend on user query. But it does not mean that this process can run without network connection.

First and continuous step is web search. This web search is doing all text-mining over all web pages that are known as web pages with science articles in one area of the research. As result of this step is indexed database of articles with most significant words in their basic form.

As the second step of the process is the creation of partial ontology. Ontology that contains classes, which represent field of research area, instances represented by words and relations

like: "*nearby*", "*the same*", "*is related*", etc. The properties of the instances will depend on frequency of occurrence. Result ontology is really only partial and contains only already known knowledge.

Third step is an ontology mapping. In general, the ontology mapping is process of finding a correspondence between concepts in different ontologies. In our case, we will search our ontologies for similarities and merge them to the one big ontology that contain not just merged knowledge, but also contain new relations, representing new interconnections, the new relationships among terms. This step requires human interaction to validate created relations, to confirm them, or change if needed. Whole process of knowledge base is shown on Fig. 1.
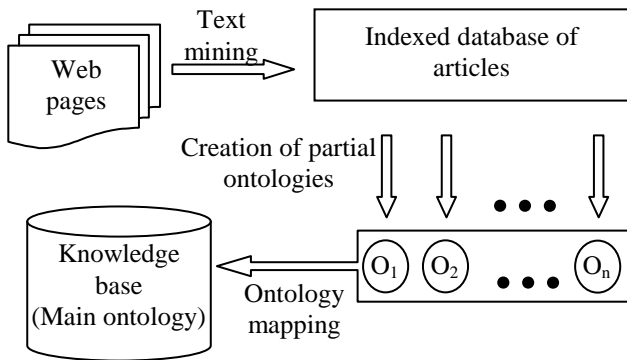
Fig. 1. Process of knowledge base creation.

This process is continuous and runs with every new article that appears on web. Beside other data, there is one property that ontology should store and that is term source documents list. It is important because we have to be able to backtrack all terms from documents. So this process keeps whole text-mining functionalities and extends them.

Query for knowledge base

When the knowledge base is completed and validated, we can start with querying. Because the source of identities origins in text-mining, queried text must pass through the same text-mining treatment as text on web sites. The query should contain also number, that will provide information of desired level of deepness. So query should look like (1).

$$q(t, d) \tag{1}$$

Where q is query, t represent term to search and d represent depth, distance of related terms.

The agent will return answers as a part of ontology and a directional graph corresponding to that ontology part. An example of such graphical representation is shown on Fig. 2.

Fig. 2. is an answer to query containing term "information" and depth "2" from test Knowledge base Ontology. You can see that "information" is often *nearby* "computer science", it is learned that *is related* to "knowledge". Also there is information that "knowledge" is *nearby* "management". Those relations are just learned from text, and not all are valid. For example term "*data*" and term "*information*" should be just *is related*, but not *the same*. Although terms are not the same, but the system correctly identifies new relationship *is related* between "data" and "computer science". Number 2 from query causes, that graph stopped at second nod from queried term.
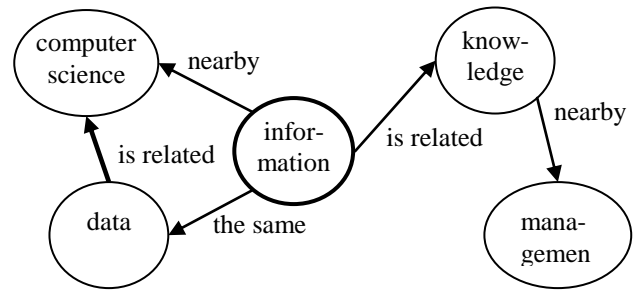
Fig. 2. Example of answer to query: *q("information", 2)*

IV. CONCLUSION

Semantic web offers great advantage for relevant search over World Wide Web, but there are still a lot of challenges that we have to defeat. I am sure, that if we combine knowledge and techniques from different sites of research, we will be able to successfully complete this journey. I hope that this work helps even a little bit to solve some problem in this kind of research. There is a long path before us, but we are here to challenge it.

REFERENCES

[1] V. DING – S. FOO, "*Ontology Research and Development. Part 2 - a review of ontology mapping and evolving*", Journal of Information Science, October 2002, vol. 28, no. 5, pp. 375-388.
[2] A. DOAN – J. MADHAVAN – R. DHAMANKAR – P. DOMINGOS – A. HALEVY, "*Learning to Match Ontologies on the Semantic Web*", The VLDB Journal, vol. 12, no. 4, pp. 303-319.
[3] N. GIIBBINS – S. HARRIS – N. SHADBOLT, "*Agent-based Semantic Web Services*", *WWW2003*, May 20–24, 2003, Budapest, Hungary.
[4] B. MATTHEWS, "*Semantic Web Technologies*", JISC Technology and Standards Watch, April 2005, www.jisc.ac.uk/uploaded_dments/jisctsw_05_02bpdf.pdf.
[5] M. SABOU – M. D'AQUIN – E. MOTTA, "*Using the Semantic Web as Background Knowledge for Ontology Mapping*", Papers from the ISWC Workshop, November 2006, pp. 1-12.

Calculating and visualization trajectories of the Chua's circuit

[1]Jozefína SCHMOTZEROVÁ, [2]Milan GUZAN, [3]Branislav SOBOTA

[1]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic
[2]Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic
[3]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]jozefina.schmotzerova@student.tuke.sk, [2]milan.guzan@tuke.sk, [3]branislav.sobota@tuke.sk

Abstract—**This paper describes Chua's circuit and one of the methods for calculating trajectories of the Chua's circuit. Program display the trajectories by using Monge projection. Change just a single parameter trajectory can be seen as a binary memory, chaos generator or oscillator of periodic waveform.**

Keywords— **attractor, chaos, Chua's circuit.**

I. INTRODUCTION

Chua's circuit shown in Fig.1 is a simple electric circuit, which has classical chaotic behavior. This chaotic behavior was presented by prof. L.O. Chua during his visit in Japan at Waseda University in 1983 [1].

Chua's circuit has found application in various fields, as confirmed by the publication in more than 1000 scientific articles [2], [3], for example in music (musical instrument
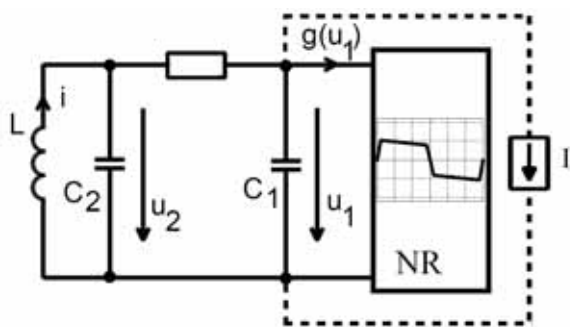
Fig. 1. Chua's circuit

ChuaSynth) [4], in cryptography [5], in health service or in hidding transmission signal [6].

Chua's circuit can generate a wide spectrum of chaotic signals - attractors (single chaos, double-scroll chaos, multi-scroll chaos) [4], [7], as well as periodic signals (single-loop cycle, multi-loop cycle) [8].

II. CALCULATING TRAJECTORIES OF CHUA'S CIRCUIT

A. Chua's circuit

That the circuit (Fig. 1) composed from standard components can display chaotic behaviour must involve

following components [1], [9]:

- One or more nonlinear elements (resistors, diodes),
- One or more resistors,
- Three or more energy-storage elements (capacitors, inductions).

The behavior of the circuit (Fig. 1) describes following system of differential equations (1) [10]:

$$C_1\left(\frac{du_1}{dt}\right) = \frac{1}{R}(u_2 - u_1) - g(u_1) - I \equiv Q_1$$

$$C_2\left(\frac{du_2}{dt}\right) = \frac{1}{R}(u_1 - u_2) + i \qquad \equiv Q_2 \qquad (1)$$

$$L\left(\frac{di}{dt}\right) = -u_2 \qquad \equiv Q_3$$

for $g(u_1)$ applies:

$$g(u_1) = m_2 u_1 + \frac{1}{2}(m_1 - m_0)(|u_1 - B_P| - |u_1 + B_P|) +$$

$$+ \frac{1}{2}(m_2 - m_1)(|u_1 - B_0| - |u_1 + B_0|) \qquad (2)$$

where:

- C_1 is capacity of the capacitor C_1 (F),
- C_2 is capacity of the capacitor C_2 (F),
- u_1 is voltage on the capacitor C_1 (V),
- u_2 is voltage on the capacitor C_2 (V),
- i is current flowing through the coil (A),
- R is resistance of resistor (Ω)

$$G = \frac{1}{R} \qquad (3)$$

- L is inductance of the coil (H),
- I is current (A),
- m_0, m_1, m_2 are slopes of IV characteristic (Fig. 2),
- B_0, B_P are values, in which IV characteristic breaks.

Program displays the chaotic attractor for the following input parameters:

C_1=0,1, C_2=2, L=0,142857, R=1,4285714, m_0=-4, m_1=-0,1, m_2=5, B_P=1, B_0=14 (4)

B. Solving of differential equations by the method Runge-Kutta

System of differential equations (1) is solved using the fourth-order Runge-Kutta method [11], [12], where:

- k_1, k_2, k_3, k_4, K are coefficients (slopes) of the method,
- h is step (interval) of the method,

TABLE I
THE FOURTH-ORDER RUNGE- KUTTA METHOD

x	y	$k_i = h.f(x, y)$
x_0	y_0	$k_1 = h.f(x_0, y_0)$
$x_0 + \dfrac{h}{2}$	$y_0 + \dfrac{k_1}{2}$	$k_2 = h.f\left(x_0 + \dfrac{h}{2}, y_0 + \dfrac{k_1}{2}\right)$
$x_0 + \dfrac{h}{2}$	$y_0 + \dfrac{k_2}{2}$	$k_3 = h.f\left(x_0 + \dfrac{h}{2}, y_0 + \dfrac{k_2}{2}\right)$
$x_0 + h$	$y_0 + k_3$	$k_4 = h.f(x_0 + h, y_0 + k_3)$
		$K = \dfrac{k_1 + 2k_2 + 2k_3 + k_4}{6}$

$$x_1 = x_0 + h \quad y_1 = y_0 + K$$

- x_0, y_0 are values of input parameters (initial conditions),
- x_1, y_1 are values of parameters in the point „1".

In the calculating have been put parameters i, u_1, and u_2 for parameters x, y. Calculating procedure for the fourth-order Runge-Kutta method is shown in the Table 1.

C. Calculating of IV characteristics

For calculating of the IV characteristics was set the voltage interval u_1=[-15V, 15V]. Coordinates of singularities 0, P$^+$
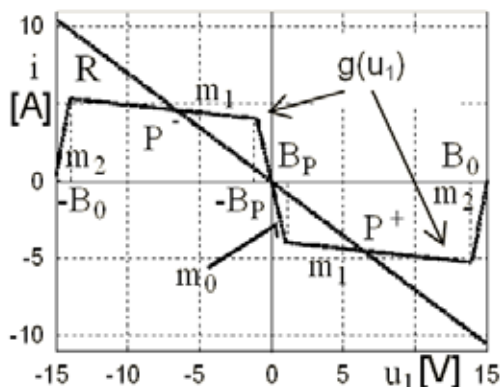
Fig. 2. IV characteristic of resistors R and NR

and P$^-$ (Fig. 2) are bound by the system (1) where:

$$Q_1=0, Q_2=0, Q_3=0 \qquad (5)$$

The resulting waveforms of currents are shown in Fig. 2.

III. VISUALIZATION TRAJECTORIES OF THE CHUA'S CIRCUIT

A. Environment

Program is written in C language, in the tool Microsoft Visual Studio 2010. The OpenGL is used for visualization.

B. Flowchart

The program is designed so that it contains only basic functions to display trajectories. These functions will be added on requirement (Fig. 3).

C. Program control

Management of the program is carried by the mouse, right click opens a context menu (Fig. 4a), wher:

- Popup menu 'Scene': allows the choice between dark (Fig. 6) ('Black') and light-coloured ('White') environment (Fig. 5).
- Popup menu 'VA characteristic': switch on ('On') (Fig.5) or switch off ('Off') the display of the IV characteristic (Fig.7).
- Popup menu 'Number of viewports': show the selected
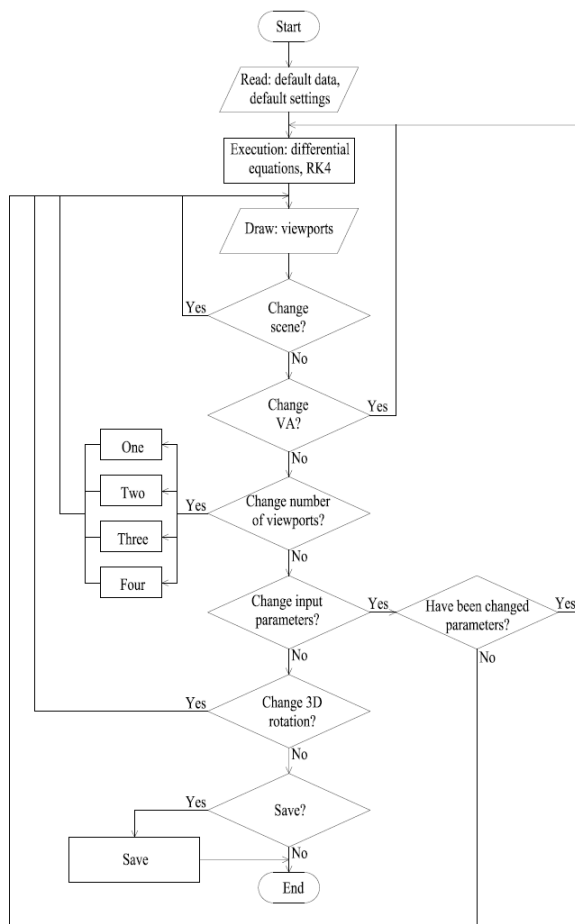
Fig. 3. Flowchart
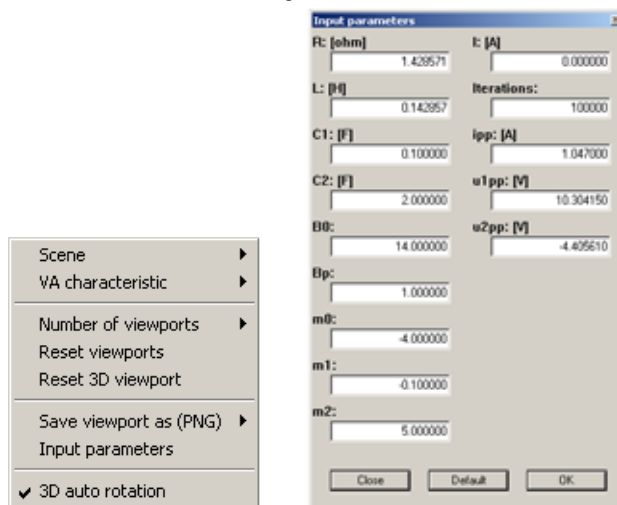
Fig. 4a. Context menu (figure on the left side)
Fig. 4b. Menu for entering of input parameters: 'Close' closed the window Input parameters, 'Default' set the default settings, 'OK' confirm the new values.
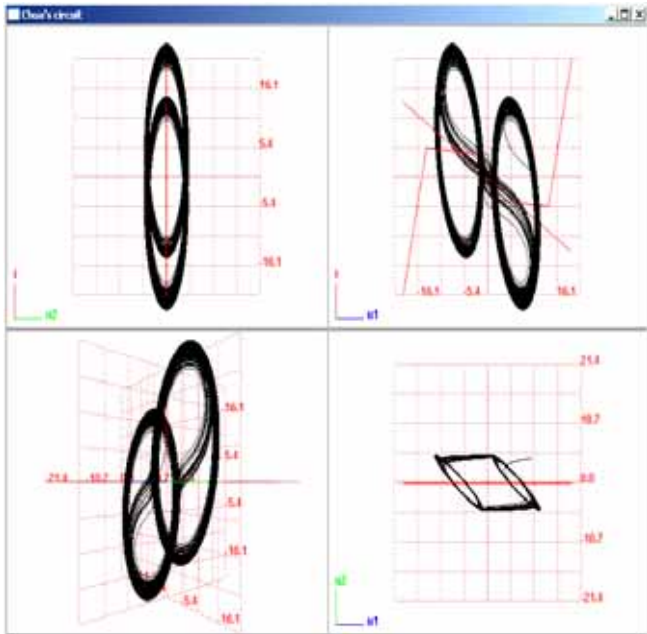
Fig. 5. IV characteristics and double-scroll chaos for parameters (4) in light-coloured environment: for switch on of IV characteristic is needed to choose the menu option: VA characteristic→On, for switch off the option: Off. Singularity points are in intersections of currents. To select a light-coloured environment is necessary to choose menu option: Scene→White.
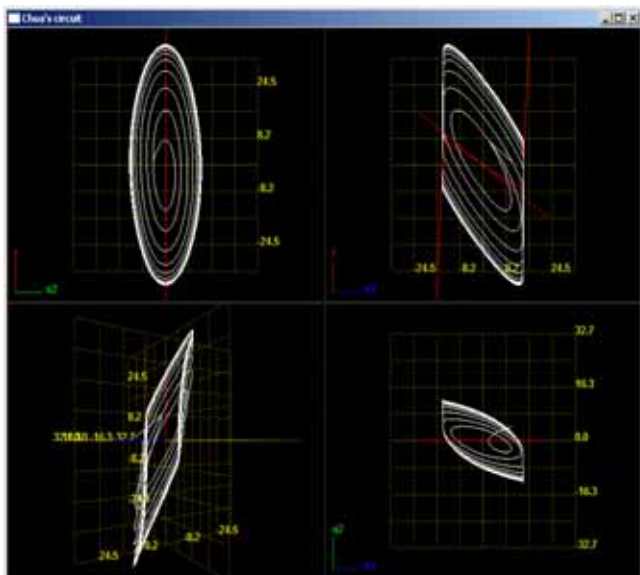
Fig. 6. Trajectory Chua's circuit in dark environment. Input parameters [10]: C_1=0,111, C_2=1, L=0,142857, R=1,4285714, m_0=-0,8, m_1=-0,5, m_2=20, B_P=1, B_0=14, i=1,047, u1=10,30415, u2=-4,40561

- number of viewports (One, Two, Three, Four).
- 'Reset viewports': after enlargement or reduction of the window, returns default view. Left click with the mouse on the center of cross-hair and current mouse movement allows to resize all viewports at the same time. Current click and move on the vertical respectively horizontal line changes the viewports only in this axis.
- 'Reset 3D viewport': after manual 3D trajectory rotation returns the initial display.
- 'Save viewport as (PNG)': prepared function for save.
- 'Input parameters': enter of input parameters for calculating of trajectory (Fig. 4b)
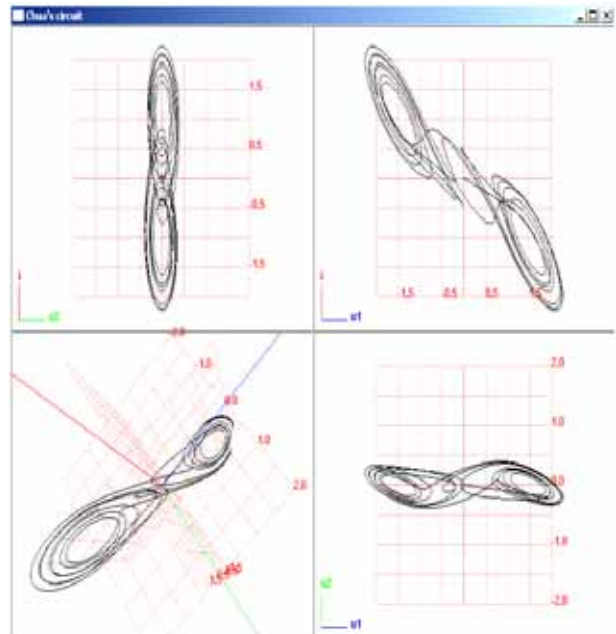
Fig. 7. Double-scroll chaotic attractor generated from Chua's circuit by changing of input parameters: i=0, u1=0, u2=0,0001, number of iteration=700. In the 3D view (on the bottom left viewport) is shown manual rotating. IV characteristics are switched off.

'3D auto rotation': default setting is checked on automatic rotation of 3D trajectory. After unmark of automatic rotation can be the trajectory rotated manually (Fig. 7). By current rotating of model (left click mouse) and pressing the key X or Y can be model rotated only around the selected axis. After selecting the menu 'Reset 3D viewport' the visualization returns in initial display.

IV. CONCLUSION

To study the properties of a chaotic signal and his widespread improvement in various fields is his calculating and visualization necessary for further analysis for example single scroll or multi-scroll chaotic attractors but also for analysis of periodic signals called periodic windows. Such visualization helps to explain the genesis of Rösler chaotic attractor and periodic window in Chua's circuit.

REFERENCES

[1] http://en.wikipedia.org/wiki/Chua%27s_circuit [cited 2011-02-15]
[2] http://scholar.google.sk/scholar?q=chua+circuit&hl=sk&as_sdt=0&as_vis=1&oi=scholart [cited 2011-02-20]
[3] http://www.eecs.berkeley.edu/~chua/publications.html [cited 2011-02-20]
[4] E. Bilotta, P. Pantano, "A Gallery of Chua Attractors," Series A vol. 61, World Scientific, 2008.
[5] T. Yang, Ch. W. Wu, L. O. Chua, "Cryptography based on Chua's circuits," Electronics Research Laboratory, College of Engineering, University of California, 1997.
[6] R. M. López-Gutiérrez, E. Rodríguez-Orozco, C. Cruz-Hernández, E. Inzunza-González, C. Posadas-Castillo, E. E. García-Guerrero, L. Cardoza-Avendaño, „Secret Communications Using Synchronized Sixth-Order Chua's Circuits", World Academy of Science, Engineering and Technology 54, 2009.
[7] Z. Elhadj, J. C. Sprott, "Generating 3-Scroll Attractors from One Chua Circuit," 2008
[8] G. Q. Zhong, F. Ayrom, "Periodicity and Chaos in Chua's Circuit," IEEE Transactions on circuits and systems, vol.. cas-32, no. 5, 1985.
[9] H. Dedieu, M. P. Kennedy, M. Hasler, „Chaos Shift Keying: Modulation and Demodulation of a Chaotic Carrier Using Self-Synchronizing Chua's Circuits," IEEE Transactions on circuits and systems – II, October 1993.
[10] M. Guzan, "Morfológia hraničnej plochy Chuaovho obvodu pri zmenách R," In: Medzinárodní konference učitelů elektrotechniky : SEKEL 2009.
[11] J. Buša, V. Pirč, Š. Schötter, "Numerické metódy, pravdepodobnosť a matematická štatistika," Košice : TU-FEI, 2006. 166 s. ISBN 80-8073-632-4
[12] http://en.wikipedia.org/wiki/Runge%E2%80%93Kutta_methods [cited 2011-02-16]
[13] M. Guzan, "Multifunkčnosť Chuaovho obvodu pri zmenách R," In: Medzinárodní konference učitelů elektrotechniky : SEKEL 2008.

Calculating the boundary surface cuts of multi-valued memory

[1]*Miroslav ZAVACKÝ,* [2]*Milan GUZAN,* [3]*Branislav SOBOTA*

[1]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic
[2]Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic
[3]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]miroslav.zavacky@student.tuke.sk, [2]milan.guzan@tuke.sk, [3]branislav.sobota@tuke.sk

Abstract— **This article deals with the calculating of an elementary memory cell, one of the non-linear dynamic systems, with the help of modern GPGPU technology. In a narrow sense, it is a guide to the alternatives of utilizing parallel principles of programming to achieve multiply faster acquisition of results.**

Keywords— **GPGPU, non-linear dynamic systems, parallel computing, state space.**

I. INTRODUCTION

Computer simulation of the processes of physical systems, as one of the means of their verification, is also very time-consuming, concerning the difficulty of computations on vast amounts of data. The use of the massive parallel computing power of GPGPU graphics cards is one of the efficient solutions of analyzing these data in state space.

II. THEORETICAL BACKGROUND OF COMPUTING

A. The elementary multi-valued memory cell

The elementary memory cell [1] of multi-valued logic (MVL) represents a dynamic system that can be implemented by two serially connected one peak of resonant tunnel diodes (Fig. 1). The I-V characteristics of the element are represented by functional dependency $i_2(u_2)$ and the I-V characteristics of the load by functional dependency $i_1(u_1)$. Both of the characteristics are piece-wice linear (PWL).
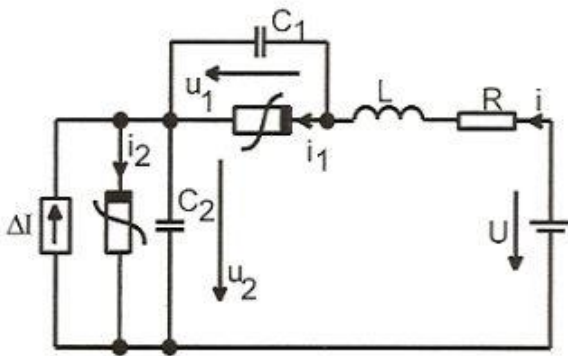
Fig. 1. The circuit of elementary MVL cell

The behavior of such structures is explainable by the use if a system of differential equations:

$$L\left(\frac{di}{dt}\right) = U - Ri - (u_1 + u_2)$$

$$C_1\left(\frac{du_1}{dt}\right) = i - f_1(u_1)$$

$$C_2\left(\frac{du_2}{dt}\right) = i - f_2(u_2) + \Delta I \qquad (1)$$

where ΔI represents current pulse for control of memory (for simplicity $\Delta I = 0$). The characteristics of non-linear components in a circuit are defined by the equation:

$$f_k(u_k) = \frac{1}{2}\left({}^k g_0 + {}^k g_3\right)u_k + \frac{1}{2}\left[\left({}^k g_1 - {}^k g_0\right)\left|u_k - {}^k U_1\right| + \right.$$
$$\left. + \left({}^k g_2 - {}^k g_1\right)\left|u_k - {}^k U_2\right| + \left({}^k g_3 - {}^k g_2\right)\left|u_k - {}^k U_3\right|\right] -$$
$$- \frac{1}{2}\left[\left({}^k g_1 - {}^k g_0\right){}^k U_1 + \left({}^k g_2 - {}^k g_1\right){}^k U_2 + \left({}^k g_3 - {}^k g_2\right){}^k U_3\right] \quad (2)$$

where ${}^k g_i$ represents conductivities and ${}^k U_i$ fracture points of I-V characteristics of a component. In a case where k=1, a given component is a load, and k=2 an element.

B. Runge-Kutta 4th order method of numerical calculations

Runge kutta 4th order is a method of iterative solution of differential equations [2], that represents an adequate compromise between fulfilling requests for low truncation error and low computing cost per step of calculation.

In our case, mentioned differencial equations of the system representing the behavior of a memory cell are in the form of:

$$y' = f(t, y) \qquad (3)$$

with an initial condition of:

$$y(t_0) = y_0 \qquad (4)$$

The formulation of the method by the given relationship is:

$$y_{n+1} = y_n + \frac{1}{6}h(k_1 + 2k_2 + 2k_3 + k_4) \qquad (5)$$

$$t_{n+1} = t_n + h \qquad (6)$$

where

$$k_1 = f(t_n, y_n) \qquad (7)$$

$$k_2 = f\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_1\right) \qquad (8)$$

$$k_3 = f\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_2\right) \qquad (9)$$

$$k_4 = f(t_n + h, y_n + hk_3) \qquad (10)$$

C. Cuts in state space

The behavior of an elementary memory cell is defined within 3-dimensional state space, whose axes are represented by the state values u_2. u_1, i. So for an arbitrary initial state given by initial conditions [u2Init, u1Init, iInit] we are able to iteratively calculate a region of attraction for an appropriate attractor with the Runge-Kutta algorithm.

The computation described further is a calculation of cuts in levels of the current axis, as shown in the picture (Fig.2).
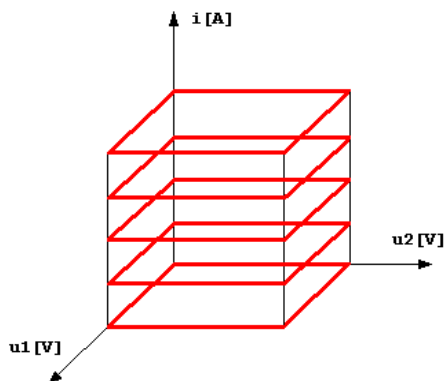
Fig. 2. Demonstration of elementary memory cell cuts in the state space

D. GPGPU

Solving tasks in state space often involves the possibility of effective use of data parallel principles in its implementation. In the case of the given system of a memory cell, the use of such principles is adequate, due to the fact that the same computational algorithm (Runge-Kutta) is applied to calculate various data (initial conditions).

GPGPU is the modern platform for parallel computing, which means using so-called graphic cards for general computing. GPGPU offers effectivity in fast processing of a large number of floating point operations while computing data extensive parallelizable problems. The latest GPGPUs incorporate several multi-processors, each including a number of processing kernels and equipped with gigabytes of memory. By comparing algorithmically equal calculations on a GPGPU and a CPU, the time demandingness of calculations on a GPGPU is tens to hundreds of time lower (depending on the extent of the problem, and optimization of the solution).

E. CUDA

CUDA (Compute Unified Device Architecture) is a parallel programming model, designed to overcome many challenges of parallel programming, offering an express opportunity to learn for programmers who are familiar with the C programming language. It is basically built on 3 abstractions: a hierarchy of thread groups, shared memory, and thread synchronization. These are exposed to the programmers as a small set of language extensions.

III. Solution analysis and implementation

A. Target platforms

The program was implemented with a goal of functioning on the NVIDIA graphic cards platform with the support of CUDA (Compute Capability version 1.1 and higher). Several restrictions closely described in section V were handled due to these versions.

As an equal choice to the model of CUDA the developers factory Khronos Group offers free standard for computing on heterogeneous platforms consisting of CPUs, GPGPUs, and other processors – the OpenCL library. For as much as both models are syntactically close, it would be able to attain platform independence by modifying the program to use the library.

B. Description of the solution

The program itself consists of two parts. The first part is represented by the primary program, which uses the "main" function. After the initial processing of input data and the memory allocation for arrays (input and output) on the host, as well as on the GPGPU device, the program continues with their initialization on the host level. Initialized arrays are further copied from host memory into the device memory. After setting the CUDA computation parameters, the program continues with executing the grid calculations in loop for the requested "i" levels. The resulting arrays are copied from the device memory into host memory and afterwards written into files. After the calculation of all chosen grids the program proceeds to free the memory and end.

The second part of the program constitutes CUDA kernel – a function which, after called from the main program, is executed in parallel on the GPGPU device. The kernel implements iterative algorithm Runge-Kutta for the trajectory calculation of the given point. As the kernel shall remain small, and its processing time shall not exceed few seconds, the problem with potentially high numbers of iterations was solved by specifying the maximum count of 1000 iterations per kernel, and by saving the values of variables dependent on loop to arrays at the end of every iteration level. These arrays are input parameters of the kernel by its cyclic calling from the main program for every value of iteration level. Thereby the consecutivity of iteration levels and thus correctness of calculations are ensured.

IV. OUTPUT AND COMPARISONS

A. Program output files

Program output is represented by the grid text files, in which the points of cut are written for given levels of current, each on new line. The format of the grid file line have a form of:

```
„u2Init   u1Init   iInit    attractor"
```

where u2Init, u1Init, iInit are initial calculation conditions (initial coordinates of given trajectory point). The attractor is an integer number that indicates the index of resulting stable singularity.

In pictures (Fig. 3 and 4), there are the figurative visualization of cuts (grid files), by the use of concerted parameters:

$$U = 0.44 \text{ V}, R = 0 \ \Omega, L = 1e-10 \text{ H},$$
$$^1g_0 = 0.1 \text{ S}, \ ^1g_1 = -0.05 \text{ S},$$
$$^1g_2 = 0 \text{ S}, \ ^1g_3 = 3.2e-2 \text{ S},$$
$$^1U_1 = 0.05 \text{ V}, \ ^1U_2 = 0.14 \text{ V}, \ ^1U_3 = 0.26 \text{ V},$$
$$^2g_0 = 8.33e-2 \text{ S}, \ ^2g_1 = -5.71e-2 \text{ S}, \quad (12)$$
$$^2g_2 = 0 \text{ S}, \ ^2g_3 = 2.81e-2 \text{ S},$$
$$^2U_1 = 0.06 \text{ V}, \ ^2U_2 = 0.13 \text{ V}, \ ^2U_3 = 0.28 \text{ V},$$
$$errorRK = 1e-3, n = 1.1, h = 1e-15,$$
$$epsI = 2e-5, epsU = 1e-3$$

The coordinates of singularities by the given system parameters:

TABLE I
SINGULARITIES COORDINATES

singularity	index	u2 [V]	u1 [V]	i [A]
S1	0	5.429320E-02	3.857068E-01	4.522618E-03
N1	1	8.621520E-02	3.537848E-01	3.501114E-03
S2	2	1.643438E-01	2.756562E-01	1.000998E-03
N2	3	3.485389E-01	9.146110E-02	2.926945E-03
S3	4	3.970883E-01	4.291170E-02	4.291170E-03

B. Comparisons of calculations time cost

This section of article deals with the comparisons of calculations for algorithmically equal programs on CPU (Intel Pentium D 2,66 GHz in configuration with 1GB RAM) and GPGPU (NVIDIA GeForce GT 8800).

Performed comparisons are shown in table (Table 2), concerning both platforms.

Graphical comparisons of calculations time consuming for various levels of current on the mentioned GPGPU are evident from the pictures (Fig. 5 and 6).

TABLE 2
TIME COST COMPARISONS OF CPU AND GPGPU

parameters	CPU [s]	GPGPU [s]	CPU / GPGPU
(12), C_1, C_2 = 2e-14 F iLevel = - 2 mA	30395	24	1266.5
(12), C_1, C_2 = 2e-14 F iLevel = 2 mA	14970	15	998
(12), C_1, C_2 = 2e-13 F iLevel = - 2 mA	153723	168	915
(12), C_1, C_2 = 2e-13 F iLevel = 2 mA	106323	133	799.4
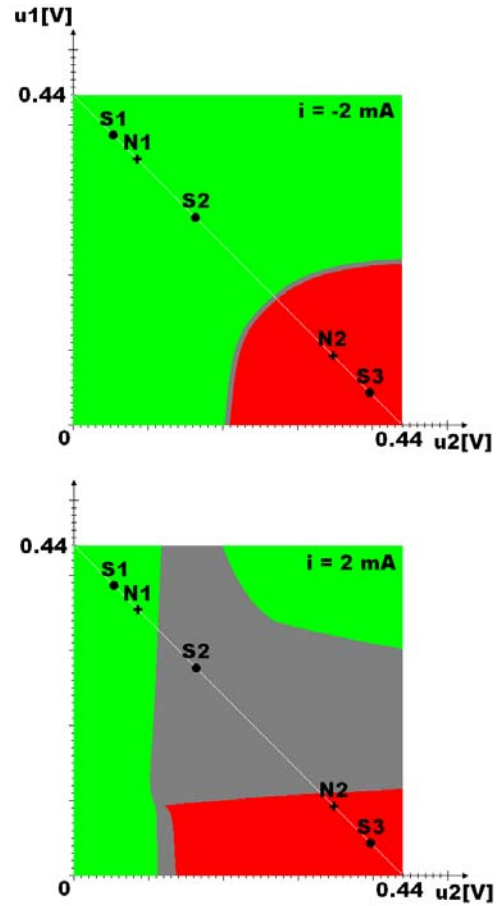
Fig. 3. Calculations of boundary surface cuts with parameters (12) and C_1, C_2 = 2e-14F. The levels of current shown in pictures.
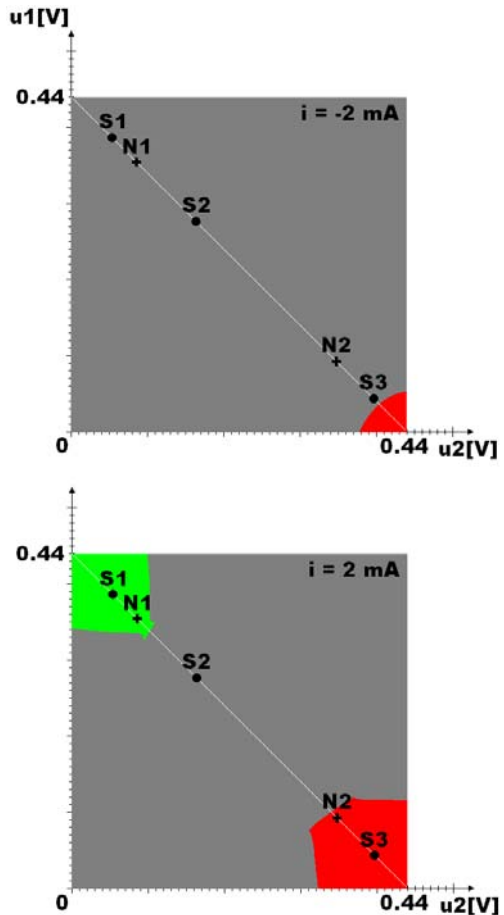
Fig. 4. Calculations of boundary surface cuts with parameters (12) and C_1, C_2 = 2e-13F. The levels of current shown in pictures.
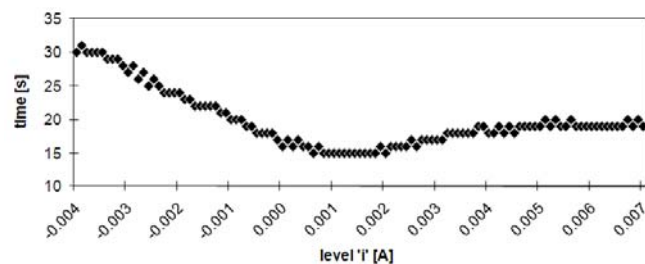
Fig. 5. Graph of time demandingness on GPGPU in dependency of iLevel
for parameters (12) and C_1, C_2 = 2e-14F
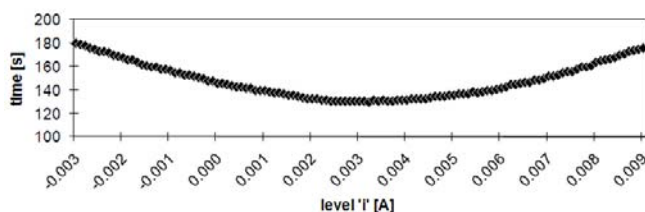
Fig. 6. Graph of time demandingness on GPGPU in dependency of iLevel
for parameters (12) and C_1, C_2 = 2e-13F

V. RESTRICTIONS

Because of Compute Capability version 1.1 the calculations in 'double' precision are not supported, therefore implementation meets only 'float' precision.

The other restriction is the calculation duration of one kernel, which shall not exceed few seconds, due to its termination without return. The description of solution to this problem is written in section III.

VI. CONCLUSION

Created program is served as a presentation of ability of dynamic systems solutions through GPGPU. Altrough it has greatly lowered the time demandingness of processing calculations comparing CPU, there still exist other possibilities to its optimalization. After modification, the program could work on modern graphic cards in the ,double' precision also.

REFERENCES

[1] M. Guzan, *Dynamic properties of multi-valued memory cell*.(Dissertation thesis). Košice, September 2001, (in slovak).
[2] D. F. Griffiths, D. J Higham, *Numerical Methods for Ordinary Differential Equations*. 1*st* Edition, 2010.
[3] [on-line] *Confession of a Speed Junkie*. Accesible on the internet <http://gpgpu-computing.blogspot.com>
[4] J. Sanders, E. Kandrot, *CUDA by example: An Introduction to General-Purpose GPU Programming*. 1*st* Edition, July 2010.

Control design of the time-delay systems

Daniel GONTKOVIČ

Department of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

daniel.gontkovic@tuke.sk

Abstract—**Systems with delays can be found in the technological process. Control design of continuous time-delay systems is being shown in this paper. This is based on Linear Matrix Inequalities (LMI) and can be utilized numerically very efficiently by using of convex optimization procedures. The aim was to provide the solution introduced by the linear matrix inequality (LMI), i.e. found LMI variables which simultaneously satisfies the limitations based on the theory of Lyapunov-Krasovskii functional. The proposed method is presented using the structural system model example.**

Keywords—**Lyapunov-Krasovskii functional, Linear matrix inequality, Time-delay systems, Time-delay, Schur complement**

I. INTRODUCTION

Delays often occur in various dynamic systems and may be a source of instability. Therefore, procedures are developed for stabilization and control systems with time delay which compensate the time-delays.

The use of Lyapunov method for the stability analysis of the time delay systems has been quite a subject of interest starting with the innovative works of Krasovskii [3]. Usually, for the stability issue of the time delay systems a modified Lyapunov-Krasovskii functionals Shaked et al. [8] can be used to achieve the delay-independent stabilization and these are applied to the controller synthesis and observer design. Progress review in this research field can be also found in Niculescu at al.[7], and the references therein.

The paper deals with the control design and stability analyzing methods based on theory of Lyapunov-Krasovskii functionals with limited speed of change of time delays on the interval $\beta = \langle 0, 1 \rangle$. The design task of the stabilizing controller for the closed-loop system is then transmuted into LMI framework and solved.

II. PROBLEM DESCRIPTION

Equations of linear dynamic systems with time delay are as follows

$$\dot{q}(t) = Aq(t) + A_d q(t-\tau) + Bu(t) \qquad (1)$$

$$y(t) = Cq(t) \qquad (2)$$

with the initial condition

$$q(\vartheta) = \varphi(\vartheta), \ \forall \vartheta \in \langle -\tau, 0 \rangle, \qquad (3)$$

where $\tau > 0$ is the delay of state, $q(t) \in I\!\!R^n$ is vector of the state, $u(t) \in I\!\!R^r$ is vector of the input variables, $y(t) \in I\!\!R^m$ is vector of the output variables, $A \in I\!\!R^{n \times n}$ is the nominal system matrix, $A_d \in I\!\!R^{n \times n}$, $B \in I\!\!R^{n \times r}$ and $C \in I\!\!R^{m \times n}$ are real matrices. Problem of the interest is to design asymptotically stable close-loop system with the linear memoryless state feedback controller of the form

$$u(t) = -Kq(t), \qquad (4)$$

where $t \geq 0$ and $K \in I\!\!R^{r \times n}$ is the controller gain matrix.

III. BASIC PRELIMINARIES

Let $Q > 0, R > 0, S$ are real matrices of appropriate dimensions and linear matrix inequality has form

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} < 0. \qquad (5)$$

Using Gauss elimination it yields

$$\begin{bmatrix} I & -SR^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} I & 0 \\ -R^{-1}S^T & I \end{bmatrix} =$$
$$= \begin{bmatrix} Q - SR^{-1}S^T & 0 \\ 0 & R \end{bmatrix} \qquad (6)$$

whereas

$$\det \begin{bmatrix} I & -SR^{-1} \\ 0 & I \end{bmatrix} = 1. \qquad (7)$$

This transformation does not change definite (5), than the next inequalities are equivalent (e.g. see [4], [5])

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} > 0 \Leftrightarrow \begin{bmatrix} Q - SR^{-1}S^T & 0 \\ 0 & R \end{bmatrix} > 0$$
$$\Updownarrow$$
$$Q - SR^{-1}S^T > 0, \quad R > 0. \qquad (8)$$

IV. STABILITY OF THE AUTONOMOUS SYSTEM

Defining of basic Lyapunov-Krasovskii functional as follows (e.g. see [1], [2], [6])

$$v(q(t), t) = q(t)^T P q(t) + \int_{t-\tau}^{t} q^T(r) Q q(r) \mathrm{d}r > 0$$
$$P > 0, \quad Q > 0. \qquad (9)$$

The result of the functional derivative of $v(q(t), t)$ is

$$\dot{v}(q(t), t) = \dot{q}^T(t) P q(t) + q^T(t) P \dot{q}(t) +$$
$$+ q^T(t) Q q(t) - q^T(t-\tau) Q q(t-\tau)(1 - \dot{\tau}(t)) < 0. \qquad (10)$$

Inserting (1) to (10) and asumming that $t \to \infty$, $\dot{\tau} \leq \beta < 1, u(t) = 0$ gives

$$(Aq(t) + A_d q(t-\tau))^T P q(t) +$$
$$+ q^T(t) P (Aq(t) + A_d q(t-\tau)) + q^T(t) Q q(t) -$$
$$- (1 - \dot{\tau}(t)) q^T(t-\tau) Q q(t-\tau) < 0. \qquad (11)$$

Creating a composite vector $q^\circ(t)$

$$q^{\circ T}(t) = \begin{bmatrix} q^T(t) & q^T(t-\tau) \end{bmatrix} \qquad (12)$$

the inequality (11) can be rewritten in a form

$$q^{\circ T}(t) \begin{bmatrix} A^T P + PA + Q & PA_d \\ A^T{}_d P & -(1-\beta)Q \end{bmatrix} q^{\circ}(t) < 0 \quad (13)$$

and can be said that the linear time delay system of form (1), (2) is asymptotically stable for delays $\dot{\tau} \leq \beta < 1$ if exists a symetric positive definite matrix P and Q such that

$$\begin{bmatrix} A^T P + PA + Q & PA_d \\ * & -(1-\beta)Q \end{bmatrix} < 0 \quad (14)$$

$$P = P^T > 0, \quad Q = Q^T > 0.$$

The problem is reduced to solving LMI i.e. to find matrices $P \in \mathbb{R}^{n \times n}$ and $Q \in \mathbb{R}^{n \times n}$, which simultaneously satisfies the boundary (14).

V. CLOSED LOOP CONTROL

The linear state feedback control law, defined in (4) gives rise to the closed-loop system as follows

$$\dot{q}(t) = (A - BK)q(t) + A_d q(t - \tau) \quad (15)$$

where $A_u = (A - BK)$ is close-loop matrix.
Then substituting (15) in (14) gives

$$\begin{bmatrix} (A - BK)^T P + P(A - BK) + Q & PA_d \\ * & -(1-\beta)Q \end{bmatrix} < 0. \quad (16)$$

Defining the congruence transform matrix

$$T_1 = T_1{}^T = \text{diag} \begin{bmatrix} P^{-1} & I \end{bmatrix}. \quad (17)$$

Premultiplying (16) from the left and right side by T_1 gives

$$\begin{bmatrix} \Delta_{11}^{\circ} + P^{-1}QP^{-1} & A_d \\ * & -(1-\beta)Q \end{bmatrix} < 0 \quad (18)$$

where $\Delta_{11}^{\circ} = P^{-1}(A - BK)^T + (A - BK)P^{-1}$.
Substituting $P^{-1} = X$ and $Y = KP^{-1}$ can be (18) rewrite such that

$$\begin{bmatrix} \Delta_{11} + XQX & A_d \\ * & -(1-\beta)Q \end{bmatrix} < 0 \quad (19)$$

where $\Delta_{11} = XA^T + AX - BY - Y^T B$.
Using the Schur complement (19) can be modified in the form

$$\begin{bmatrix} \Delta_{11} & X & A_d \\ * & -Q^{-1} & 0 \\ * & * & -(1-\beta)Q \end{bmatrix} < 0. \quad (20)$$

Defining the second congruence transforms the matrix as follows

$$T_2 = T_2{}^T = \text{diag} \begin{bmatrix} I & I & Q^{-1} \end{bmatrix} \quad (21)$$

and then premultiplying from the right and left side by (21) modifying it in the form

$$\begin{bmatrix} \Delta_{11} & X & A_d Q^{-1} \\ * & -Q^{-1} & 0 \\ * & * & -(1-\beta)Q^{-1} \end{bmatrix} < 0. \quad (22)$$

Substitution $Q^{-1} = Z$ leads to inequality

$$\begin{bmatrix} XA^T + AX - BY - Y^T B & X & A_d Z \\ * & -Z & 0 \\ * & * & -(1-\beta)Z \end{bmatrix} < 0. \quad (23)$$

For system (1), (2) the necessary and sufficient condition for the stable nominal control (4) is that there exist positive definite symmetric matrices X, Z and matrix Y, such that the following LMIs are satisfied

$$\begin{bmatrix} XA^T + AX - BY - Y^T B & X & A_d Z \\ * & -Z & 0 \\ * & * & -(1-\beta)Z \end{bmatrix} < 0$$

$$X > 0, \quad Z > 0. \quad (24)$$

The control law gain matrix is then given as

$$K = YX^{-1}. \quad (25)$$

VI. ILLUSTRATIVE EXAMPLE

To demonstrate the algorithm properties it was assumed that system is given by (1), (2), where

$$A = \begin{bmatrix} -2.6 & 0.0 & 0.8 \\ -1.2 & 0.2 & 0.0 \\ 0.0 & 0.5 & -3.0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix},$$

$$A_d = \begin{bmatrix} 0.00 & 0.02 & 0.00 \\ 0.00 & 0.00 & -1.00 \\ -0.02 & 0.00 & 0.00 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 0 \\ 7 & 1 \\ 1 & 0 \end{bmatrix}.$$

Solving (24) for LMI matrix variables $X > 0$, $Z > 0$ and Y using SeDuMi package for Matlab, given task was feasible with

$$X = \begin{bmatrix} 0.4241 & 0.0024 & 0.0071 \\ 0.0024 & 0.4434 & 0.0274 \\ 0.0071 & 0.0274 & 0.3010 \end{bmatrix},$$

$$Z = \begin{bmatrix} 0.9017 & 0.0003 & 0.0007 \\ 0.0003 & 0.9049 & 0.0076 \\ 0.0007 & 0.0076 & 0.6938 \end{bmatrix},$$

$$Y = \begin{bmatrix} -0.0875 & 0.3548 & -0.1473 \\ -0.2402 & 0.0080 & 0.1537 \end{bmatrix},$$

where $P^{-1} = X$ and $Q^{-1} = Z$. The controller gain matrix K_1 can be obtained from the relation (25) for $\beta = 0.1$, where

$$K_1 = \begin{bmatrix} 0.2016 & 0.8360 & -0.5607 \\ -0.5751 & -0.0113 & 0.5255 \end{bmatrix},$$

$$\rho(A_u) = \{-1.4823, -2.3516 + 0.6367i, -2.3516 - 0.6367i\}.$$

It is evident that the eigenvalues spectrum $\rho(A_u)$ of the closed control loop is stable.
Setting $\beta = 0.5$ the gain matrix K_2 takes the form

$$K_2 = \begin{bmatrix} -0.2290 & 0.9657 & -0.6317 \\ -0.6009 & 0.0368 & 0.5390 \end{bmatrix}.$$

$$\text{eig}(A_u) = \{-1.4530, -2.5236 + 0.7793i, -2.5236 - 0.7793i\}.$$

and the spectrum of the eigenvalues $\rho(A_u)$ of the closed control loop is stable, too.
Solving with respect to the controller gain matrix K_3 and $\beta = 0.99$ results

$$K_3 = \begin{bmatrix} -0.1347 & 1.0606 & 0.3224 \\ -0.5949 & 0.1755 & -0.0115 \end{bmatrix},$$

Even for this case the eigenvalues spectrum $\rho(A_u)$ of the closed control loop is stable, since

$$\rho(A_u) = \{-1.4823, -2.3516 + 0.6367i, -2.3516 - 0.6367i\}.$$
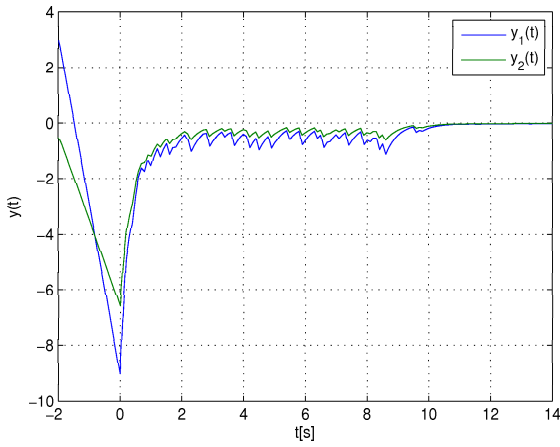
Fig. 1.   Response of the closed-loop system for $\beta = 0.1$
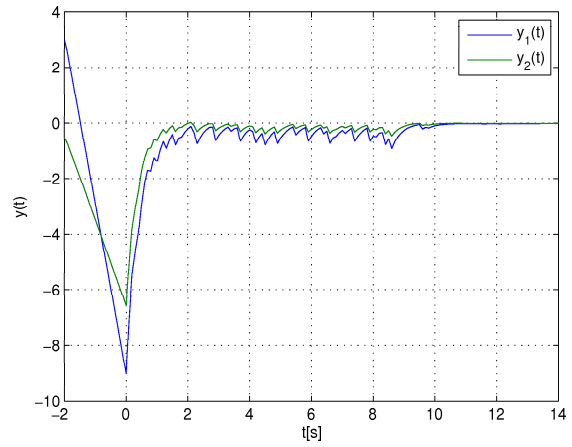
Fig. 3.   Response of the closed-loop system for $\beta = 0.99$
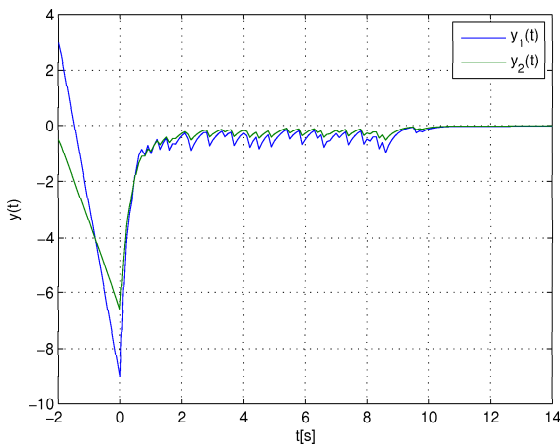
Fig. 2.   Response of the closed-loop system for $\beta = 0.5$
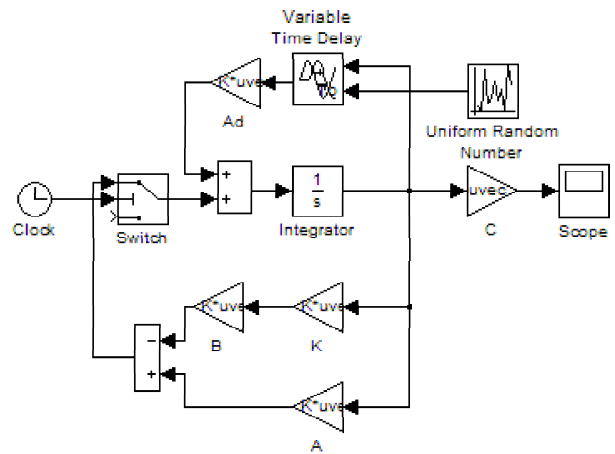
Fig. 4.   Response of the closed-loop system for $\beta = 0.99$

VII. SIMULATION

Time-delay linear dynamic system (1), (2) is controlled into the steady state with nonzero initial conditions $[-1, 0.5, 3]^T$ for the above solutions.

In Figure 1 to 3 time response of the closed-loop system with time-varying delay is shown. Solving the issue of stability for the system with time delay is secured only to delay $0 < \dot{\tau} \leq 1$. Any proposed control response system in the negative time is unstable, because one's eigenvalue A_d is unstable. System reaches the steady state under $11s$.

A simulation scheme of a closed-loop system with time delay is shown in Figure 4.

VIII. CONCLUSION

The method uses the standard LMI numerical optimization procedures to manipulate the system feedback gain matrix as the direct design variable. It is necessary to propose an asymptotically stable closed-loop system.

To solve the problem of stability in an independent delay today we typically use Lyapunov-Krasovskii functionals with limited speed of change of time delays on the interval $\beta = \langle 0, 1 \rangle$ as shown above.

Finally the design method is illustrated by an example.

REFERENCES

[1] FILASOVÁ, A. - KROKAVEC, D.: Asymptotically stable control design for time-delay systems. In *Proceedings of the 17th International Conference Process Control '09, June 9-12, 2009, Štrbské Pleso*, Slovak Republic, STU Bratislava, Bratislava, Slovak Republic.

[2] FILASOVÁ, A. - KROKAVEC, D.: Global asymptotically stable control design for time-delay systems. AT&P Journal Plus, No. 2, 2009.

[3] KRASOVSKII, N. N.: Stability of motion: Application of Lyapunov's Second Method to Differential Systems and Equations with Delay. Standford: Stanford University Press, 1963.

[4] KROKAVEC, D., FILASOVÁ, A.: Optimal Stochastic Systems. Košice: ELFA, s.r.o., 2002. (in Slovak)

[5] KROKAVEC, D. - FILASOVÁ, A.: *Discrete-time Systems*, Elfa, Košice, 2008. (in Slovak)

[6] KROKAVEC, D., GONTKOVIČ, D.: Output feedback control design of time-delayed systems. In *Proceedings of the 9th International Scientific-Technical Conference Process Control 2010*. June 07-10, 2010, Kouty nad Desnou, Czech Republic, University of Pardubice, Czech republik, 2010.

[7] NICULESCU, S.I., et al.: Stability and robust stability of time-delay systems: A guided tour. In *Stability and Control of Time-delay Systems*. Springer-Verlag, Berlin, 1998.

[8] SHAKED, U., et al.: Bounded real criteria for linear time systems with state-delay. *IEEE Transactions on Automatic Control*, 43, 1116–1121, 1998.

Creation of the process map using
ARIS Business Architect and QPR ProcessGuide

[1]*Rastislav HOŠÁK,* [2] *Peter KUBIČKO,* [3] *Roman MIHAĽ*

[1,2,3] Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]rastislav.hosak@tuke.sk, [2]peter.kubicko@tuke.sk, [3]roman.mihal@tuke.sk

Abstract—**This article discusses the standards of modeling business processes. Analyzes the various standards and points to the missing parts, which were in general application, and clearly determined how would models of business processes look like. In the second part of this paper work with two modeling software is described - ARIS Business Architect and QPR ProcessGuide. The possibilities that are offered by this tools are described there. Modeling with them is mutually compared and in the end there is the recommendation concerning which instruments you can choose to create models of business processes.**

Keywords—**process, process map, object, model, ARIS Business Architect, QPR ProcessGuide.**

I. INTRODUCTION

Process modeling is a very discussed topic today. There are few specialists and lack of specific instructions on exactly how a procedural map should look. Norms and standards for this area gradually emerge. The international standard norm describing the modeling processes in details still has not been released.

II. STANDARDS FOR PROCESS MODELING

The process is always modeled as a structure of mutually interwoven activities. Depending on the clarity of modeling each activity can be described as a separate process. Important activities are not random, but only on the basis of defined initiatives.

Impulses may be external or internal. Under external impulses we understand those which come from the environment of the process and are defined as events. When talking about internal impulses, these are the impulses coming from inside the model, and therefore they are called the state of process.

Activities are linked to each other. The summary of these activities creates a process, which is defined by its structure. Continuity of individual operations is described by linkages. Links are defined by different types of activity patterns in the process, from a simple sequence to a parallelism.

It is generally known that different types of models and norms are defined for the modeling of business processes. Overview of standards in the modeling of business processes is brought by Institute CIMOSA Association (TABLE 1).

TABLE I
OVERVIEW OF STANDARDS FOR MODELING BUSINESS PROCESSES CIMOS

ISO 14258 Terms and rules for modeling organizations		
ISO 15704 Requirements for the reference architecture methodologies		
Frameworks	Languages	Modules
CEN/ISO 19439 Frameworks for modeling	CEN/ISO 19440 Constructors for modeling	ENV 13550 Service to create a model EMEIS
ISO 15745 Framework of integration and application	ISO 18629 Process Specification Language	ISO 15531 Manufacturing data exchange
ISO 15288 Life Cycle Management	ISO/IEC 15414 ODP Language for description of organizations	ISO DIS 16100 Software for support of production
	BPMI / BPML Language for modeling business process	IEC/ISO 62264 Integration of control systems
	OMG/RfP UML business process description	

A. Business Process Modeling Notation (BPMN)

BPMN is designed for the model specification. The comprehensibility of the graphical part is given by the specification. The main purpose is to present and describe the processes understandably. There are three kinds of processes in the BPMN:

- Private processes - dealing with the internal processes of the organization.
- Public abstract processes - designed to gather information outside the private processes.
- Processes of cooperation - describe cooperation between two or more business elements.

Execution mechanism for BPMN is Business Process Modeling Language (BPML) [1]. It is a modeling language for process description, which is based on Extensible Markup Language (XML). Version of modeling language BPML first appeared in 2002. The concept focuses mainly on the collaboration among trading partners, which is also its basis. BPML is generally defined as an executive language, which

means that the models can be run in a defined environment.

The basic element of the language is an activity that is associated with a function. Activities are organized hierarchically. Business process can be viewed as a special type of activity that contains other embedded processes. Activities are divided into simple and compound. Simple activities are those, which cannot be divided into other activities. In contrast, compound activities are those which are composed of more activities. Compound activity is also referred to as a set of activities in some literatures.

When dealing with the context, that is characterized as a general behavior of all activities. Some special fields define the context as cooperating activities which use it in order to exchange information using functions defined in the properties.

Process in BPML is considered to be a complex activity that contains its own context for launching activities. Processes are divided into internal, exceptional or compensational. Internal can be understood as those that are run in a certain context. The exceptional processes are designed for handling of exceptions and errors. Compensatory processes allow the process to return to the state before starting the main process. An important feature of these processes is that they can be run only once after the main process is ended.

Plans are a series of specific time events in which the process is running. Plans form a timetable of currently running processes. They may represents time limit of the process duration.

B. Unified Modeling Language (UML)

UML is a visual object-oriented methodology for modeling of large systems [15]. It was primarily designed to unify different approaches during the establishment of the specific requirements in the process of creating a software product. However, it became a universal standard language for creating documentation of any systems. The language specification contains a number of diagrams describing the modeled system from all the possible previews.

UML provides structures for capturing the system from both, static and dynamic point of views. The creators of the systems thus are enabled to create unified designs that can be effectively shared and communicated across multiple teams or other tools and enables their uniform interpretation.

The first version of UML (UML 0.8) was created in 1995. Today the version of UML 2.0 is used and it is used largely in the design of systems.

C. IDEF standards

We see the concept IDEF (Integrated DEFinition methods), as a group of methods designed for a complex enterprise architecture modeling support. IDEF is a result from the research program of ICAM. Nowadays, there are six methods developed in the level of practical use; these are the methods ranging from IDEF0 to IDEF5. Another eight methods are under development (from IDEF6 to IDEF14).

D. ISO standards

ISO standards for the issue can be divided into three basic groups.

ISO 14 258

It specifies the basic terms and rules for the computer models for the standardizing of the business process modeling. It focuses mainly on the definition of the basic concepts in the modeling of the business. It defines the system life cycle and its basic phases, concepts, hierarchy, structure and behavior. It determines the requirements for other standards in order to achieve the universality of the model.

ISO 15 704

The standard defines basic and general requirements for the business models, the reference architecture and the associated methodology.

The standard emphasizes the general approach, but also a clear definition of, mainly, the project objectives and the scope of his place. It is necessary to work with business processes, human factors and not least with the various supporting technologies in the methodology based on this standard.

ISO 18 629

The standard defines the modeling language and semantics of the processes. It also specifies the neutral definition of processes. Usability of data is determined and used throughout the whole product life cycle, ranging from the production planning to the actual production of a particular product, in other words, managing the production of the product.

Unlike previous standards, ISO standard 18 629 is currently under the development and only the first part of the standard is completed.

III.   SOFTWARE FOR WORKING WITH BUSINESS PROCESSES

There are many programs on the market for the work with the business process models (process map). They differ mainly in the development environment. Each producer of such software has its own standards as how its application should look like, and what ergonomics it should offer. This is generally true not only about this concrete field, but applies to software in general.

Particular applications offer different options. Each allows you to create the process maps; this results from the determination of the software product. Many from them also offer other options such as simulations, creation of reports, graphs, based on simulation. They allow adding of various attributes to particular procedural steps.

It should be noted that business process models have not a single graphical representation yet. This means that each element is represented by different graphical symbols in different software. For this reason, the same model created in one modeling tools can look much differently in the second.

As already mentioned above, there is lots of software of this type, therefore we focus only on two. Specifically the software ARIS Business Architect and QPR ProcessGuide.

A. ARIS Business Architect

ARIS Business Architect is part of the ARIS Design platform. It represents a standard for the enterprise-wide process design. ARIS Design Platform products help to solve various business problems such as: who does what, in what order to do various actions, what has been achieved, and so on. These processes identify the basic structural, organizational and technical problems and their solutions.

ARIS Business Architect is a web-based application; its applicability is primarily for business process management.

B. QPR ProcessGuide

QPR Collaborative Management Suite integrates process management of organization. It allows you to evaluate existing systems and to build a company-wide process management system. QPR ProcessGuide is its part.

QPR ProcessGuide is an intuitive tool offering a quick creation of process maps [3]. It offers modeling, simulation and simple web publishing. It also provides a wide range of electronic documentation system according to ISO 9001:2000. At the same time, it provides access to a wide range of users.

IV. COMPARISON OF MODELING SOFTWARE

The process of acquiring customers was chosen as the object of the modeling. This process may be implemented in three ways.

The first way is represented by the object "Acquisition of an address of a potential customer" and the object "Address of a potential customer". These objects are linked to the CRM
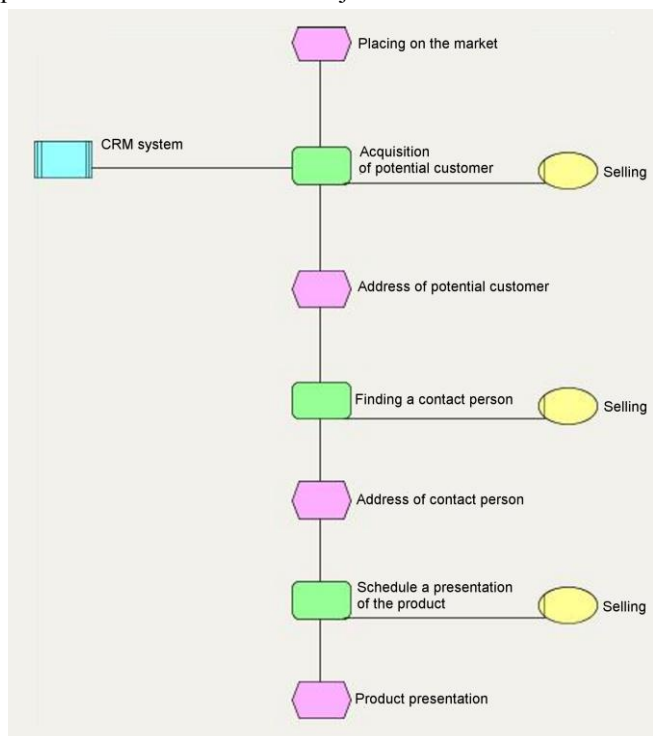
Fig. 1. Process of acquiring customers – ARIS Business Architect

system, which represent the technical equipment for the support of the process. After obtaining the addresses, the sale is realized, as can be clearly seen from the model.

The second way of how to obtain customers is described in the following two objects. The first object is "Finding a contact person" and the second is "Address of a contact person". This method is used with the first case and is one of the alternatives. After obtaining the contact person, the sale of the product is realized.

The third way of how to obtain a customer is the product presentation. The activity is represented by the last two objects, "Projection of the presentation of a product" and "The product presentation". After completing the presentation, the sale of the product is realized.

The model was created in both mentioned applications.

ARIS Business Architect, similarly to QPR ProcessGuide, can be considered to be interactive software for planning and implementation of the business processes. Both programming devices operate on the same principle of the modeling of the business processes.

Figure 1 shows the supply chain model created in as "EPC diagram" (Event Process Chain diagram) in ARIS Business Architect.

ARIS Business Architect offers a wider selection of tools and different types of models that can be created [2]. We selected, "Fast Model", "Administrative process" and "EPC diagram".

Creation of a process map is sequential, the "fast model" being created in the first step. Thank to this model, the person who creates the model specifies his idea of the resultant model. The essential properties of the objects are modeled there. The system does not differentiate either functions, nor events in this type of modeling. All elements in the model are identified as objects.
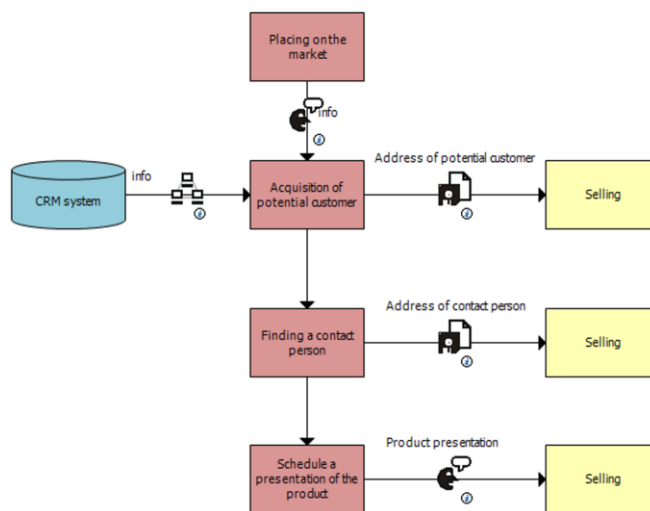
Fig. 2. Process of acquiring customers – QPR ProcessGuide

"EPC diagram" was created on the basis of the "Administrative process". In this case, objects are differentiated, according to their type, into events, functions, computer, organizational unit and the links between them.

The principle of the modeling is the same as in the previous type of modeling of the fast model. The importance of individual objects did not change, but they are represented by new objects. As noted above, the objects have received a new specification and they are not labeled as objects any more. The first type of the object is the function. The functions can be defined as technical tasks or activities designed to support the achievement of the target. In this case, the functions define the act that should be done in order for the process to be successful. Another type of object that is defined in the administrative process is an event. Events represent the important conditions of the business process.

EPC diagram is created by copying the administrative process directly into the EPC model. EPC diagram is the final product and is used to manage the consolidation and to improve the states of the corporation.

Figure 2. shows the same process, but ismodeled in the program QPR ProcessGuide.

QPR ProcessGuide has also its own template with different

types of objects. We used the procedural steps called "Activity" in the modeling. They represent a further sub-activity for which a specific system or a person is responsible. The links between the activities have their own names in order to make it clear what the particular flow represents. Links are defined by their type which divides them into information, control or material flow. CRM system is represented by the object "Information store".

We can see that the models are similar; however, they are not identical. EPC diagram modeled in ARIS Business Architect has several objects. This is because the objects also represented information (such as a customer address). A person who follows that model or a person who creates information systems realizes that the object is important and will not overlook it. On the other hand, the objects are represented only as flows in the model from QPR ProcessGuide. There is a greater chance that the person overlooks something important already during the process of modeling. The output of the precession step is represented only by flow. It is represented by the rotation of objects activity - output in the EPC diagram.

Model in QPR ProcessGuide can be understood as directed, in the case of simulation, weighted. But it is not standardized. It also offers drilling down in the sub processes which enhances the clarity of the model. In the other case, complicated models might be created. They would be hardly comprehended. In addition, the user has the option to view only those objects in the model which he is currently interested in.

As already mentioned, ARIS Business Architect allows you to create several types of models. EPC diagram, which can be created in this software, can be regarded as a standard due to the fact that this software is so widely used. Modeling in this program is intuitive and offers a wide range of possibilities, this firstly slowing down the work in it.

V. CONCLUSION

Process modeling is an important activity in designing information systems. Process map is needed to obtain the quality certificates ISO. Although the modeling of business processes is embedded in the standards and frameworks, there is no exact description of objects and how to process map should look like.

From this perspective, it is only up to a user which software he will choose for the modeling.

It should be noted that modeling tools are not uniform. There is no single modeling environment. Generally valid graphic symbols and descriptions for individual objects do not exist. There already exist rules and standards dealing with this issue, but they miss a single directive, which would define the form of the process modeling.

ACKNOWLEDGMENT

REFERENCES

[1] V. Řepa, "Podnikové procesy. Procesní řízení a modelováni. 2," updated and expanded edition, Grada Publishing, Praha, 2007, ISBN 978-80-247-2252-8.

[2] A-W. Scheer, "ARIS – od podnikových procesu k aplikačním systémum," Brno 1999, ISBN 80-238-4719-8.

[3] QPR software Plc, "QPR ProcessGuide," User's Guide, version 8.1.1.

[4] N.G. Leveson, M.P.E. Heimdahl, H. Hildreth, J.D. Reese, "Requirements Specification for Process-Control Systems," IEEE Transactions on Software Engineering, vol. 20, no. 9, pp. 684-707, 1994, doi:10.1109/32.317428.

[5] Y. Wu, J. Chen, X. Ba, "Application and Implementation of SMS Platform in Project Management Information System Approval Workflow," wcse, vol. 1, pp.77-80, 2010 Second WRI World Congress on Software Engineering, 2010.

[6] L. Ladryová, C. Irgens, "Process knowledge generation and knowledge management ti support product quality in the process industry by supervisory control and data acquisition (SCADA) open systems," Production Planning&Control, Vol. 17, No. 2, March 2006, pp. 94-98, C Taylor&Francis, ISSN 1366-5871.

[7] P. Kubičko, "Orchestrálna partitúra nielen pre riadenie ITIL/ITSM procesov - Procesný model," In: Orchestrácia a automatizácia ITIL/ITSM procesov, international Workshop day, Bratislava, 2010.

[8] P. Budaj, P. Kubičko, L. Lukáč, "Cyklické procesné riadenie v sociálnych službách," In: Disputationes Scientificae Universitatis Chatholicae in Ružomberok, Ružomberok: KU 2010, Vol. 10, No.4, pp. 88-94, ISSN 1335-9185

[9] M. Hrnčiar, "Možnosti priemetu výsledkov výskumných projektov do vzdelávania v oblasti kvality, In: Kvalita pedagogiky kvality : vedecké práce z 1. vedeckej konferencie, Alexander Dubček University of Trenčín, 2004, pp. 31-33, ISBN 80-88914-58-2.

[10] J. Koziorek, "Education of industrial control systems at VSB-Technical University of Ostrava," Czech Republic. In CONTI'2008 - The 8th International Conference on Technical Informatics, 5-6 June 2008, Timisoara, Romania. Vol.2, pp. 135-140. ISSN 1844-539X

[11] L. Landryová, P. Osadník, M. Valas, O. Winkler, " Integrating tools and methodologies in FutureSME delivery platform", Seminar ASR '2010 "Instruments and Control" C 2010, VŠB-TUO, Ostrava, ISBN 978-80-248-2191-7.

[12] D. Popescu, Q. Meng, A.A. Ionete, C. Ionete, R. Aguridan "Remote vs. Simulated, Virtual or Real-Time Automation Laboratory" in International Conference on Automation and Logistics, Shenyang, China, 2009, pp. 1410 – 1415, ISBN 978-1-4244-4794-7.

[13] L. Révészová, D. Paľová, "Základy modelovania podnikových procesov," 1. Edition, Technical University of Košice, , 2009, ISBN 978-80-553-0174-7.

[14] M. Hrnčiar, "Analýza procesov poskytovanie služieb s využitím prístupov ISO/IEC 15504," In: Diagnostika podniku, controlling a logistika III, International Scientific Conference, University of Žilina, 2006, pp. 81-84, ISBN 80-8070-527-5.

[15] G. Booch, J. Rumbaugh, I. Jacobson, "Unified Modeling Language User Guide," The (2nd Edition) (Addison-Wesley Object Technology Series), Addison-Wesley Professional 2005, ISBN0321267974.

Data Mining and Machine Learning Intrusion Detection Techniques in Encrypted Internet Communication

Michal AUGUSTÍN

Dept. of Computer and Informatics, FEI TU of Košice, Slovak Republic

michal.augutin@tuke.sk

Abstract — **The emerging number of threads and attacks gives recency to Intrusion Detection and Prevention Systems (IDPS). These threads are even present in encrypted communication (e.g. HTTPS), which have an impact on availability. This paper presents intrusion detection system based on data mining and machine learning supervised algorithm C4.5 on collected and analyzed packet headers. The proposed system can classify normal network activities and main attack types such as DoS and Probe.**

Keywords — **intrusion detection, packet headers, anomaly intrusion detection, supervised machine learning algorithms**

I. INTRODUCTION

Nowadays network security is a very important field of Informatics and with the emergence of Internet a lot of sensitive information, such as financial transactions are sent through this medium. This gave the need to protect this content – the use of encryption technologies and protocols. Some of the most widely used communication protocols that use encryption are SSL/TSL, IPsec, SSH etc [1] [2] [3].

Computer attacks, intruders, hackers and other malicious programs, viruses try to gain control over these systems, crash these systems down (e.g. DoS) or commit other frauds. Stable and secure encryption protocols with correct implementation partially removed these risks, and so it became harder for attackers to break encrypted communication, nevertheless other attacks in encrypted communication are still possible.

Consequently, computer networks have become more prone to different types of network attacks. Intrusion Detection Systems (IDS) offer methods to protect these networks of many attacks, some IDS use signature based rules other use anomaly detection. Signature based detection works well on known attack patterns which are stored in a database or a file. These IDS must be frequently updated in order to keep their signatures up to date for efficient detection of novel attacks or threats. Signature-based detection suffers from the inability to detect new type or zero-day attacks, those attacks that have not signatures included in databases. Another considerable aspect is the continuous growth of the signature database in size.

Anomaly intrusion detection techniques can be applied to identify novel or zero-day attacks against computers and network infrastructure [4] [5]. Anomaly detection algorithms generally work in two phases; a training phase and a detection phase. Generally the training phase contains benign traffic and algorithms learn and create models out of it. In the detection phase the current traffic is compared with the learned profiles/models and any abnormality from the normal behavior is marked as anomaly.

It is also a challenge to detect new worms and viruses in early stages of spreading, an example is a single worm W32/SQLSlam-A which infected 75 000 machines in 30minutes and caused failures of major network services [7].

Another frequent attack is Denial of Service (DoS) or Distributed Denial of Service (DDoS), which causes significant damages on information availability to intended users. There are two forms of DoS attacks, one form causes services to crash and other floods services [8] Some of these software tools for performing DoS attacks are available online, such as hping, httping, evil-ping etc [10] [11] .

Some of the top intrusion attacks are listed in Table I available from McAfee Labs [9].

TABLE I
SOME TOP INTRUSION DETECTIONS ATTACKS

Data	Tools	Category
HTTP Flood	Net Tools 5	Denial-of-Service
Jping	Jping.c	Denial-of-Service
exe.file.b.ident		WORM-General

Encryption communication protocols encrypt content, this content is then sent over the Internet through the TCP/IP protocol, the data is then "broken" into packets for routing purposes (encapsulation) with IP-header and IP-data. IP headers can not be encrypted, because they are used for routing.

Threads and attacks are also a common case against encrypted communication, an example is requests to generate several thousand or million page on HTTPS server – SSL over HTTP.

Collecting, analyzing and storing these IP headers and applying data mining on large amount of IP headers, as well as using machine learning techniques can prevent attacks even on encrypted communication. This paper presents a data mining and machine learning combination approach on detecting intrusive attacks with the use of an open source network intrusion prevention and detection system SNORT and mySQL[6].

II. COLLECTING INFORMATION FROM HEADERS

Information stored in packet header is extracted using SNORT IDS to capture all relevant header information (not packet payload – payload is encrypted and so irrelevant) and then stored locally in a mySQL database. Information from packets such as IP header, TCP header, UDP header and ICMP header ale collected. Collected data features are IP Source and IP Destination, IP Version, IP Header length etc. Fig. 1 shows the table data structure of collected information in mySQL database. Table II shows some examples of data obtained from SNORT

Fig. 1. Shows table structure of stored information in mySQL database through open source SNORT IDS

TABLE II
SOME FEATURES PROCESSED

Feature Description	Data Type
Number of TCP packets	Integer
Number of TCP source port	Integer
Number of TCP destination port	Integer
Number of UDP packets	Integer
Number of UDP source port	Integer
Number of UDP destination port	Integer
Number of ICMP packets	Integer

III. APPLYING DATA MINING ON COLLECTED HEADERS

Data mining is a common process of extracting patterns from a large date sets by combining different methods of statistics and artificial intelligence with databases. The primary reason for using data mining is to assist in the analysis of collections of a given observation and behavior.

The pre-processing needs to collect a large data set, and then to uncover the patterns presented in the data. The training set was used to apply data mining algorithms and set a normal behavior of the network. The task of data mining was to correctly classify benign and abnormal activity on the network. The domains of interest were **number of packet headers**, **source** of the packet and **time**.

This gave the ability to cluster data and discover group structure in the data that is in some way similar.

For data mining purposes RapidMiner, an open-source system, was used with an access to database mySQL [12]. RapidMiner graphical user interface was used to execute a SQL query and plotting features were used to vizualise the given results.

SELECT DISTINCT(iphdr.ip_src), count(iphdr.ip_src) AS count, sum(iphdr.ip_len) AS length FROM iphdr LEFT JOIN event ON iphdr.cid = event.cid WHERE iphdr.ip_src <> '"+ip_src+"' AND event.timestamp = '"+fromDateTime+"' GROUP BY iphdr.ip_src;

SQL query counts the number of a given IP source and a sum of IP length (of a given IP) with respect to time. Result of data mining were plotted on a graph. With a benign activity the structure of the data had regular patterns as shown in Fig.
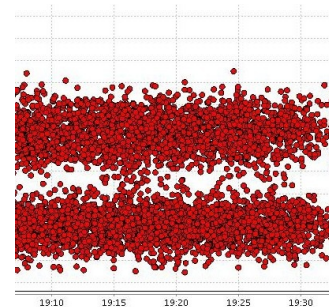
Fig. 2. Shows sum of packet header length (of a given IP) Y ax in respect to time X ax in network traffic flow

2.

To evaluate the ability to detect an attack, a DoS attack on a Local Area Network (LAN) was performed and data collected after this attack was data mined. Fig. 3 shows a significant change in the IP length of packets coming from the same IP source address – peaks in the graph.
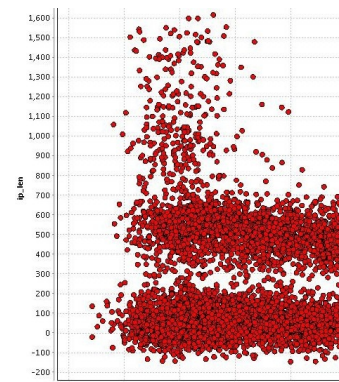
Fig. 3. Shows sum of header length (of a given IP) Y ax in respect to time X ax in network traffic flow. Peaks in the graph indicate malicious activity.

IV. APPLYING MACHINE LEARNING C4.5 ALGORITHM ON HEADERS

Machine Learning, a branch of artificial intelligence, is concerned with the design and development of algorithms that allow computers to evolve behaviors based on data – to learn [13].

Algorithm types in machine learning are organized into taxonomy, based on the desired outcome – Supervised Learning generate a function that maps inputs to desired outputs.

Some well known algorithms of supervised machine learning are those that create decision trees for example **ID3** (Iterative Dichotomizer 3) by Quinlan or an extension **C4.5** used for classification [14]. Decision tree approach considers with features of packet headers gained from network traffic. Decision tree algorithms have a high performance in classifying unknown attacks [15].

Decision tree approach considers with features of packet headers gained from network traffic. Decision tree consists of non-terminal nodes (roots, internal nodes) and terminal nodes (leaves) which efficiently classify data.

Root node is the first attribute with test conditions to split every record toward each internal node. At first decision tree is trained with known data by a learning type, afterward it can classify new data. This algorithm can predict new data by starting from root to internal nodes and then to a leaf node consisting of class – each non-terminal node has to be tested to meet given conditions.

C4.5 builds its decision from the training data by using the concept of information entropy [13]. Fig. 4 shows the process of creating a C4.5 decision tree and the tested attributes in an open-source machine learning program RapidMiner [12].

V. BAGGING

Bagging is a method for improving results of machine
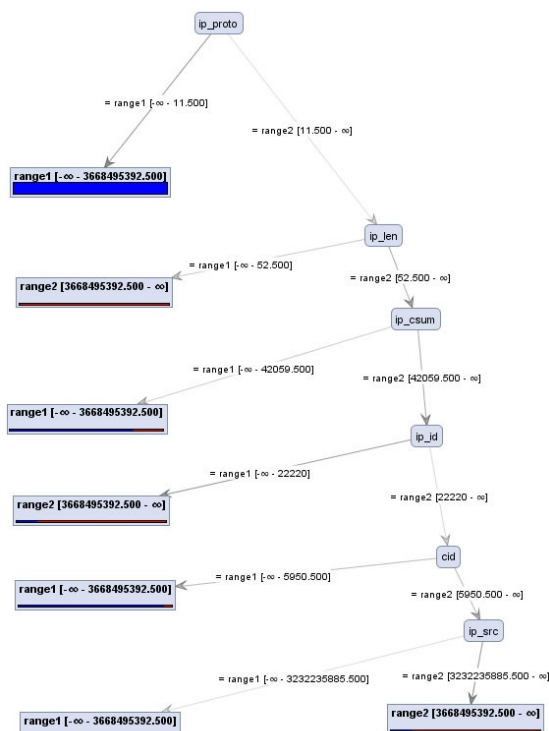
Fig. 3. C4.5 decision tree

learning classification algorithms. This method was formulated by Leo Breiman and its name was deduced from the phrase "bootstrap aggregating" [16] [17]. The bagging method creates a sequence of classifiers H_m, m=1,…,M in respect to modifications of the training set. These classifiers are combined into a compound classifier, this can be interpreted as a voting procedure.

VI. CONCLUSION

This paper proposed various ways of detecting intrusions using data-mining and machine learning supervised algorithms – decision trees C4.5 with an efficient data processing features extracted from packet headers. Packet headers contain important information that should be considered even in encrypted communication protocols, such as SSL and other, for better attack detection on networks. Detection rates can be improved by combining data-mining techniques and decision trees – Algorithm C4.5 and thus creating composite detection. Bagging methods can be also applied to improve results classification algorithms of machine learning.

REFERENCES

[1] Protecting Transmission Control Protocol/ Internet Protocol (TCP/IP) Networks: NIST Guidance, http://www.itl.nist.gov/lab/bulletns/bltnapr06.htm, 12-12-2010
[2] Internet Protocol (IP) Encryption Solutions, http://www.certesnetworks.com/securitysolutions/ip-encryption.html, 10-12-2010
[3] RFC 4252 – The Secure Shell (SHH) Authentication Protocol, http://tools.ietf.org/html/rfc4252, 03-03-2010
[4] Zero Day Exploits: The Holy Grail, http://netsecurity.about.com/od/newsandeditorial1/a/aazeroday.htm, 05-09-2010
[5] Anomaly detection can prevent network attacks – Techworld.com, http://features.techworld.com/networking/2338/anomaly-detection-can-prevent-network-attacks/, 12-11-2010
[6] Snort, About Snort, http://www.snort.org/snort, 03-01-2011
[7] W32/SQL Slam-A – A Win32 worm, http://www.sophos.com/security/analyses/viruses-and-spyware/w32sqlslama.html, 07-08-2010
[8] CERT/CC Denial-of-Service, http://www.cert.org/tech_tips/denial_of_service.html, 10-09-2010
[9] Top Intrusion Attacks, http://www.mcafee.com/threat-intelligence/ips/top-attacks.aspx, 14-02-2011
[10] How to defend against DdoS attacks, http://www.computerworld.com/s/article/94014/How_to_defend_against_DDoS_attacks, 14-03-2011
[11] Evil Ping – Free Download, http://www.supershareware.com/evil-ping-free/, 03-03-2011
[12] Rapid – I, http://rapid-i.com/content/view/181/190/lang,en/, 05-02-2011
[13] Kristína Machová, Strojové učenie Princípy a algoritmy, ISBN 80-89066-51-8, elfa 2002
[14] Kristína Machová, Strojové učenie v systémoch spracovania informácií, ISBN 978-80-8086-130-8, elfa 2009
[15] Network Intrusion Detection and Classification with Decision Tree and Rule Based Approaches, P. Sangkatsanee, N. Wattanapongsakorn, and C. Charnsripinyo, IEEE Inter Conference: ISCIT, 2009
[16] Breiman, L.: Bagging predictors. Technical Report 421, Department of Statistics, University of California at Berkeley, 1994
[17] Friedman, http://www-stat.stanford.edu/people/faculty/friedman.html, 12-02-2011

Design and Simulation of Control for Nonlinear Model Ball&Beam

Peter ŠUSTER

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

peter.suster@tuke.sk

Abstract — **This paper deals design of control algorithms using approximate linearization method, namely PD and state control algorithm for nonlinear simulation model Ball&Beam. In this paper is given a mathematical – physical model whereby was created nonlinear simulation model Ball&Beam in the Simulink environment. The proposed control algorithms are implemented into control structures with the aim tracking desired position of ball along beam. The programmed simulation schemes of the control structures for tracking reference trajectory are verified in the Matlab/Simulink language.**

Keywords — **nonlinear model, pole – placement method, PD control algorithm, state control algorithm.**

I. INTRODUCTION

Model Ball&Beam is the typical example of a natural nonstable system, which is mostly used in teaching, based on its properties, in the validation and testing of various proposed control algorithms in control structures [2].

In this paper are designed PD and state control algorithm for control of the simulation nonlinear model, which is based on mathematical – physical model Ball&Beam. The proposed control algorithms are implemented into simulation scheme in Maltab/Simulink language with the aim to track desired trajectory, which represents the position of ball along beam.

The control algorithms based on linear model in input – output or state description were used for comparison with algorithms, which are designed by nonlinear method of synthesis of exact linearization in my work to the dissertation exam for model Ball&Beam [8].

II. MATHEMATICAL – PHYSICAL MODEL BALL&BEAM

The model Ball&Beam (B&B) (Fig.1) is nonlinear dynamic SISO model with one input (u_{MU} – voltage (machine units) that control servomotor) and one output (r – position of ball along beam).
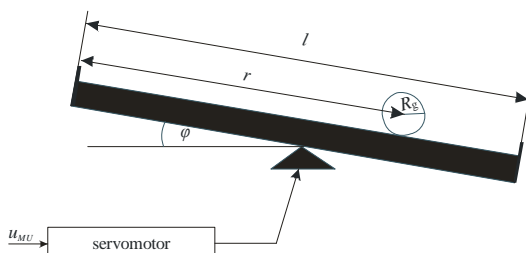
Fig. 1 Ball&Beam model

Model B&B is a simplification of Ball&Plate (B&P) model, which is located at the Laboratory of Cybernetics at the Department of Cybernetics and Artificial Intelligence [1].

The model B&B is divided into two subsystems namely subsystem *Movement of ball along beam* and subsystem *Servomotor* (Fig.2). Sensor converts position ball along beam in meters to position in Machine Units (MU).
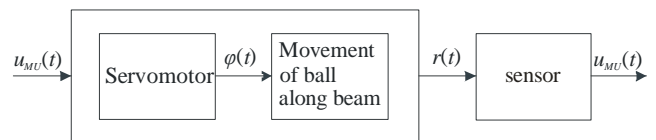
Fig. 2 Ball&Beam – subsystems

In compiling of the mathematical – physical model of subsystem *Movement of ball along beam* is based on the Euler - Lagrange equations in the shape

$$\frac{d}{dt}\left\{\frac{\partial}{\partial \dot{q}_i(t)} L(t)\right\} - \frac{\partial}{\partial q_i(t)}\{L(t)\} = \frac{\partial}{\partial q_i}\{W(t)\} \qquad (1)$$

where $L(t) \equiv L = E_k$ - E_p is Lagrangian, $q_i(t)$ is i – th generalized parameter and $W(t)$ is work, that undertaken by the nonconservative forces. The result after determination of total kinematic E_k and potential energy E_p and defining generalized parameters ($q_1(t) = r(t)$ a $q_2(t) = \varphi(t)$) are two nonlinear differential equations

$$\left[m_g + \frac{J_g}{R_g^2}\right] \cdot \ddot{r}(t) - m_g \cdot r(t) \cdot \dot{\varphi}^2(t) + m_g \cdot g \cdot \sin(\varphi) = 0 \qquad (2)$$

$$\left[J_t + m_g \cdot r^2(t)\right] \cdot \ddot{\varphi}(t) + m_g \cdot g \cdot r(t) \cdot \cos(\varphi) = F_t \cdot l \cdot \cos(\varphi) \qquad (3)$$

where $r(t)$ is current position of the ball, $\varphi(t)$ is angle of the rotation of the beam, F_t is force acting in direction of the beam, m_g is mass of the ball, R_g is radius of the ball, J_g is moment of inertia of the ball, J_t is moment of inertia of the beam and g is acceleration of gravity. The equation (2) expresses impact angle of the beam on the movement of the ball and equation (3) expresses impact ball on the angle of the beam, which neglects. Programmed simulation scheme of the subsystem *Movement of ball along beam* is in Fig.3.
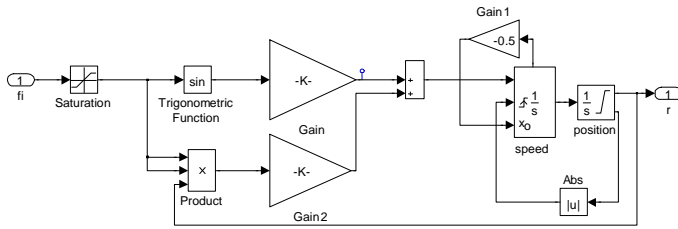
Fig. 3 Simulation scheme of the subsystem of Movement of ball along beam

A complete description of the model B&B is still necessary to define subsystem *Servomotor*, which can be described nonlinear differential equation in the shape

$$\frac{d\varphi(t)}{dt} = \omega.sign[u_{MU}(t) - u_\varphi(t)] \, pre \, [u_{MU}(t) - u_\varphi(t)] \geq sens \quad (4)$$

resp.

$$\frac{d\varphi(t)}{dt} = 0 \, pre \, [u_{MU}(t) - u_\varphi(t)] < sens \quad (5)$$

where ω is nominal angular velocity of the servomotor and *sens* is characteristic for nonlinear element type three – position relay. Simulation scheme of the subsystem *Servomotor* (Fig.4) was programmed from nonlinear differential equations (4) and (5). [1], [2]
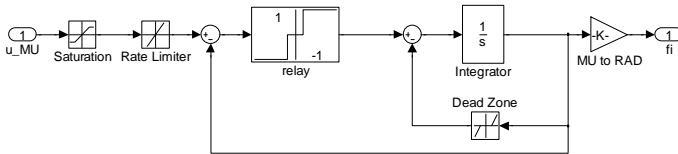
Fig. 4 Simulation scheme of the subsystem of Servomotor

III. DESING PD CONTROL ALGORITHM

For proposal control algorithms using approximate linearization method is necessary linearized nonlinear model B&B in point [0,0]. In this case is necessary linearized both subsystems *Movement of ball along beam* (2) and *Servomotor* (4), (5).

After treatment will have nonlinear differential equation (2) shape

$$\ddot{r}(t) = K_m \sin \varphi(t) \quad (6)$$

Whereas the rolling of the beam is limited in range ± 5% for linearization of equation (6) is used fact that $\sin \varphi \cong \varphi$ for the small angles, then transfer function of the subsystem *Movement of ball along beam* is in the shape

$$\ddot{r}(t) = K_m \varphi(t) \quad \rightarrow \quad F_m = \frac{Y(s)}{U(s)} = \frac{K_m}{s^2} \quad (7)$$

Subsystem *Servomotor* (Fig.4) contains hard nonlinearities and therefore can not linearized equation (4) and (5) using Taylor series. Therefore method of gradual integration was used to obtain a linear approximation of the subsystem *Servomotor* [4].

After applying algorithm of the identification method was obtained linear approximation of the subsystem Servomotor in the shape

$$F_s = \frac{K_s}{T_s s + 1} \quad (8)$$

The sensor with static characteristic, where the sensor's constant is K_{sn}, is used to detect position of the ball along beam. The resulting transfer function of the model B&B is then in the shape

$$F = F_{sn} F_m F_s = \frac{K_{sn} K_m K_s}{s^2 (T_s s + 1)} = \frac{K_{B\&B}}{s^2 (T_s s + 1)} \quad (9)$$

The obtained linearized model of input – output description (9) is used to design the PD control algorithm with aim to track the given trajectory, that includes control on the equilibrium [0,0] and also control on the steady state.

The aim is design the control algorithm, which ensures that control error converge to zero during tracking reference trajectory i.e. $e(t) \rightarrow 0$.

Based on the resulting transfer function of the model B&B (9) was proposed the parameters for PD control algorithm using programmed algorithm of method of standard shapes of Graham – Lantrop for consider control law in the form

$$u(t) = K(e(t) + T_d \frac{de(t)}{dt}) \quad (10)$$

The proposed simulation scheme of closed – loop system is in Fig.5, where is implemented the proposed PD control algorithm to control nonlinear model B&B.
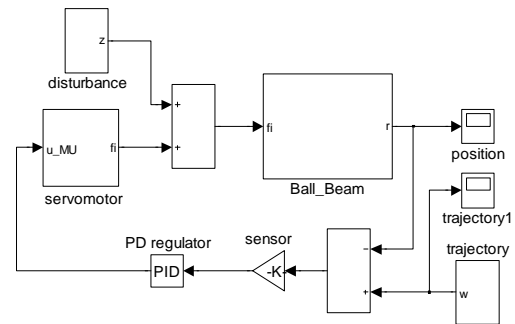
Fig. 5 Simulation scheme of feedback control structure using the PD control for the model B&B

The resulting graph of the tracking reference trajectory using PD control algorithm is in Fig. 6, from which is seen that output of the nonlinear model B&B tracks reference trajectory. The oscillations around steady states are caused to the fact that the proposed control is applied to the nonlinear model.
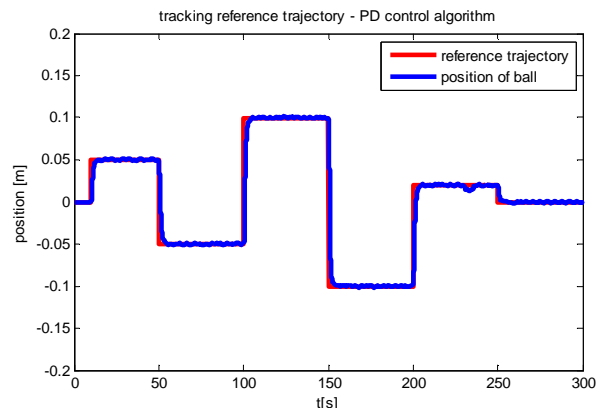
Fig. 6 Tracking reference trajectory – PD control algorithm

IV. DESING STATE CONTROL ALGORITHM

The another proposal control algorithm using approximate linearization method is state control algorithm, which requires a linear model B&B (9) in state space

$$x(t) = Ax(t) + bu(t)$$
$$y(t) = c^T x(t)$$
(11)

where state matrix of the model B&B are consider in the shape

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & K_{B\&B} \\ 0 & 0 & -\dfrac{1}{T_s} \end{bmatrix}, \; b = \begin{bmatrix} 0 \\ 0 \\ \dfrac{1}{T_s} \end{bmatrix}, \; c^T = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$
(12)

Base on the matrix A, b, c^T of the state description of the model B&B, suitable chosen roots $p = [p_1, p_2, p_3]$ and using function of Control Toolbox *place* based on pole – placement method is designed vector of gains K for consider control law in the form

$$u(t) = -Kx(t) + Nw(t)$$
(13)

The gain N of the control algorithm, which ensures tracking trajectory is determined from the relation [7]

$$N = \frac{1}{c^T (I - (A - bK))^{-1} b}$$
(14)

The proposed control algorithm ensures control on the equilibrium but it not sufficient for control on the steady state, therefore the chosen roots p were optimized using programmed genetic algorithm yet, that was achieved the desired behavior of the nonlinear simulation model B&B. [9]

The proposed simulation scheme using the designed state control for nonlinear model B&B in the feedforward control structure is in Fig.7.
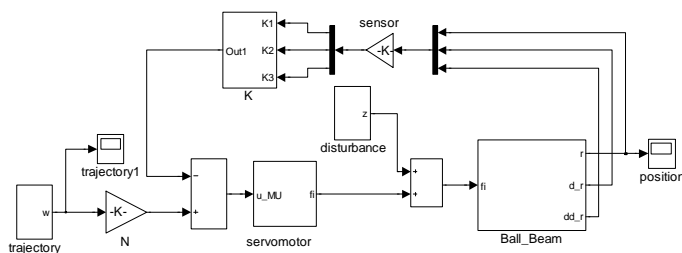
Fig. 7 Simulation scheme of control structure with forward regulator using state control algorithm

The resulting graph of tracking reference trajectory of model B&B using state control is in Fig.8. The reference trajectory is the same as when was used PD control algorithm. The Fig.8 shows that output of the model B&B tracks the desired trajectory but when compared with graph of using PD control (Fig.6) are visible the larger oscillation around the steady state.

The proposed PD and state control algorithm are verified on the nonlinear simulation model B&B, which created conditions for their further use for real model Ball&Plate. The model B&P is divided into two subsystems namely subsystem for a axis x and subsystem for a axis y, while model B&B represent one of the subsystems its simpler variant of the model B&P. [1]
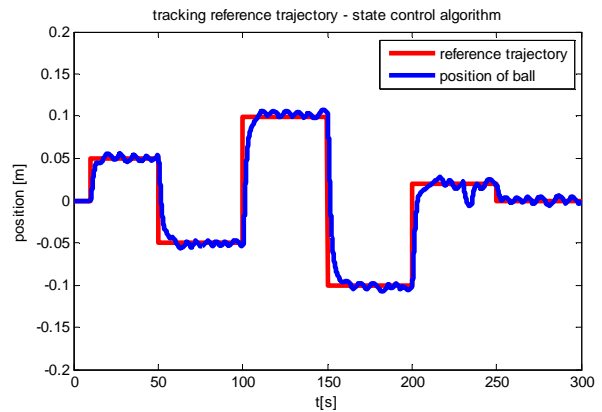
Fig. 8 Tracking reference trajectory – state control algorithm

V. CONCLUSION

The paper presents the design of PD and state control algorithm for nonlinear simulation model B&B. The approximate linearization of the nonlinear model is used in the both proposals and based on the obtained linear approximation of system is designed control for nonlinear model. These proposed control algorithms will serve as standard for comparison with the proposed algorithms using nonlinear methods of synthesis, namely exact linearization, gain – scheduling method and control based on the Ljapunov function, for model B&B resp. B&P in my future dissertation work, where the main aim is create a software tool for design of control algorithms for nonlinear physical systems.

REFERENCES

[1] Humusoft Praha : CE 151 Education Manual Ball & Plate
[2] P. Vojčinák, M. Pieš, R. Hájovský : Design of LQR control for physical model B&B, In Technical Computing Prague 2009: Proceedings of 17th Annual Conference, 19. 11. 2009, Praha: Kongresové centrum ČVUT, 2009, ISBN 978-80-7080-733-0; (in Czech)
[3] L. Madarasz : Theory of automatization control (Linear and nonlinear systems), ALFA Bratislava, 1990, 333 s., ISBN 80-05-00550-4 (in Slovak)
[4] M. Fikar – J. Mikles : Identification of systems, STU Bratislava, 1999. (in Slovak)
[5] O. Modrlák : Synthesis of regulation circuit. Study materials [online]. Liberec: TU, 2004. aviable on the internet : http://www.fm.vslib.cz/~krtsub/fm/modrlak/pdf/tar1_syn.pdf (in Czech)
[6] J. Sarnovský – A. Jadlovská – P. Kica : Theory of optimal and adaptive systems. Košice: Elfa, 2005, ISBN 80-8086-020-3 (in Slovak)
[7] D. Krokavec and A. Filasová : *Discrete-Time Systems*, Elfa, Košice, 2008. (in Slovak)
[8] P. Šuster : Contribution to the analysis and synthesis of nonlinear dynamic systems, Thesis for dissertation exam, TU Košice, Faculty of Electrical Engineering and Informatics, Košice, Slovakia, 2010. (in Slovak)
[9] I. Sekaj: Evolution computing and their using in practice, Bratislava - IRIS, 2005, 157 s., ISBN 80-89018-87-7 (in Slovak)

Detection of Abnormal Sounds in Urban Environment

Eva VOZÁRIKOVÁ

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

eva.vozarikova@tuke.sk

Abstract—**This paper describes the detection of abnormal acoustic events such as gun shot and breaking glass in urban environment. These sounds are not typical in this surroundings, they have rarely occurence and represent potential danger. For this detection task Hidden Markov Models (HMMs) were used as a classifier. Accoustic signal is representing by MFCC and the selected set of MPEG-7 descriptors such as AudioSpectrumCentroid, AudioSpectrumSpread and AudioSpectrumFlatness.**

Keywords—**Acoustic event, MFCC, MPEG-7**

I. INTRODUCTION

In the future, surveillance systems will be a common part of any public or private security systems. The complex surveillance system [1] can be created by the fusion of audio and video information. The effectiveness of the surveillance system depends on the environmental conditions. The visual part of the detection system will probably fail in terms of the bad light condition, or if the problematic situation is not in the visual field of the surveillance camera. On the other hand, the audio part of system is very sensitive on the similarity of sounds. In generally, extreme weather conditions limit the performance of any surveillance system. Surveillance systems are usually used at monitor public places [2], stadiums, vehicles and stations of public transport [3], etc. Some dangerous situations are more easily detectable via audio information [4] than the visual information e.g. calling for help, sounds of gun shots, etc. The goal of each surveillance system is help to protect life and property. Detection systems should generate alerts only if dangerous events are detected. Therefore, it is very important to reduce false alarm as a result of incorrectly classified pattern. The proposed detection system is created to recognize potentially dangerous situations via sound information. Especially, our effort is to detect two types of acoustic events such as gun shot and breaking glass. These sounds represent abnormal behavior and they point to existence of some dangerous situation e.g. a robbery, etc.

Acoustic signals have information redundancy and for that reason is necessary to specify effective feature extraction methods. The effective feature extraction should highlight the relevant information and reduce the number of input data by removing irrelevant information. Several methods [5] can be used to represent the acoustic signal, e.g. inspired by speech parametrization like MFCC [6], [7], or inspired by descriptors of MPEG-7 standard [8], etc. MPEG-7 is a standard, focused on the describing of the multimedia content. It is oriented on the indexing, searching and retrieval of audio using the 17 low level descriptors [2], [9], [10]. These descriptors can capture the nature of input acoustic signal. This paper is focused on the

feature extraction method that include the advantage of speech parametrization like MFCC and the selected set of MPEG-7 low level descriptors. In classification stage HMMs are used. The influence of the intensity of background sounds to the proposed detection system is investigated.

II. FEATURE EXTRACTION METHODOLOGY

The efficient feature extraction is a very important phase of the detection process, because the recognition performance directly depends on the quality of the extracted feature vectors.

A. Mel-Frequency Cepstral coefficients - MFCC

MFCC belongs to a most popular feature extraction methods. It is usually used to represent the characteristics of speech signals. MFCC coefficients are computed following way: a segment of signal is divided into short frames, where the parameters of the signal are constant. The Hamming window method was applied on the frames. Then, they are transformed to the frequency domain via the Fast Fourier Transform (FFT), and then the magnitude spectrum is passed through a bank of triangular shaped filters. The energy output from each filter is then log-compressed and transformed to the cepstral domain via the Discrete Cosine Transform (DCT) [8], [11]. The relation between the Mel-frequency and the frequency is given by the Eq.(1):

$$Mel(f) = 2595 \times log_{10}\left(1 + \frac{f}{700}\right), \qquad (1)$$

where f is frequency in Hertz.

B. MPEG-7 low level descriptors

MPEG-7 is an ISO/IEC standard developed in by the Moving Picture Experts Group (MPEG). MPEG-7 became an international standard in September 2001 [10]. It includes the part dealing with audio information i.e. MPEG-7 Audio. There are defined 17 low level descriptors, see Fig.1.

In our experiments basic spectral descriptors namely AudioSpectrumCentroid, AudioSpectrumSpread and AudioSpectrumFlatness were chosen according to the good results that were presented in the works [2], [5], [8].

AudioSpectrumCentroid - ASC
The audio spectrum centroid (ASC) [10] gives the centre of gravity of a log-frequency power spectrum. All power coefficients below 62.5 Hz are summed and represented by a single coefficient, in order to prevent a non-zero DC component

Fig. 1.   MPEG-7 low level descriptors.

Fig. 2.   Architecture of proposed system.
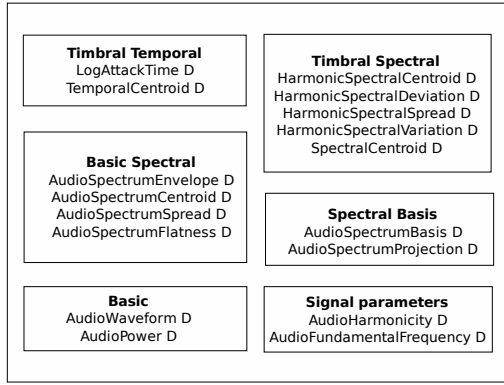
and / or very low-frequency components which can have a disproportionate weight. For a given frame of signal, ASC descriptor is computed from the modified power coefficients and their frequencies. In the Eq. (2), $P'(k')$ represents the power spectrum and $f'(k')$ represent corresponding frequencies.

$$ASC = \frac{\sum_{k'=0}^{(N_{FT}/2)-K_{low}} log_2\left(\frac{f'(k')}{1000}\right) P'(k')}{\sum_{k'=0}^{(N_{FT}/2)-K_{low}} P'(k')}. \qquad (2)$$

The ASC gives information on the shape of the power spectrum. It indicates whether in a power spectrum are dominated by low or high frequencies and can be regarded as an approximation of the perceptual sharpness of the signal.

AudioSpectrumSpread - ASS
The audio spectrum spread (ASS) [10] is also called instantaneous bandwidth. It is measure of the spectral shape. In MPEG-7, it is defined as the second central moment of the log-frequency spectrum. For a given signal frame ASS is computed following way:

$$ASS = \frac{\sum_{k'=0}^{(N_{FT}/2)-K_{low}} \left[log_2\left(\frac{f'(k')}{1000}\right) - ASC\right]^2 P'(k')}{\sum_{k'=0}^{(N_{FT}/2)-K_{low}} P'(k')}. \qquad (3)$$

ASS descriptor is extracted by taking the root-mean-square (RMS) deviation of the spectrum from its centroid ASC. The ASS gives indications about how the spectrum is distributed around its centroid. A low ASS value means that the spectrum can be concentrated around the centroid, whereas a high value reflects a distribution of power across a wider range of frequencies.

AudioSpectrumFlatness - ASF
The audio spectrum flatness (ASF) [10] reflects the flatness properties of the power spectrum. More precisely, for a given signal frame, it consists of a series of values, each one expressing the deviation of the signals power spectrum from a flat shape inside a predefined frequency band. In MPEG-7, the power coefficients are computed from non-overlapping frames where the spectrum B is divided into 1/4 octave resolution

logarithmically spaced overlapping frequency bands. For each band b, a spectral flatness is estimated as the ratio between the geometric mean and the arithmetic mean of the spectral power coefficients within this band:

$$ASF(b) = \frac{\sqrt[hiK'_b-loK'_b+1]{\Pi_{k'=loK'_b}^{hiK'_b} P_g(k')}}{\frac{1}{hiK'_b - loK'_b + 1} \sum_{k'=loK'_b}^{hiK'_b} P_g(k')}, \qquad (1 \le b \le B).$$
$$(4)$$

For all bands under the edge of 1 kHz, the power coefficients are averaged in the normal way. For all bands above 1 kHz, power coefficients are grouped $P_g(k')$. The terms $hiK'b$ and $loK'b$ represent the high and low limit for band b. High values of ASF coefficients reflect noisiness, on the other hand, low values indicate a harmonic structure of the spectrum.

III. SYSTEM OVERVIEW

Our proposed system was developed for the purpose of abnormal sound detection such as gun shot and breaking glass, see Fig. 2. The well known HTK Toolkit [12] was used for training of the Hidden Markov Models (HMMs). For each acoustic event and also for background sounds HMMs were trained up to 64 Probability Density Functions (PDF). As a feature vector the speech based features MFCC with log-energy coefficient _E and three descriptors of MPEG-7 such as ASS, ASC, ASF were used.

MFCC were computed from the signal divided into the 30 ms frames with 50% overlapping between the neighboring frames. 29 triangular band filters were used. MFCC feature extraction algorithm generated 13 coefficients (12 static coefficients and 1 log-energy coefficient _E). Extraction of MFCC was done by HTK tool HCopy.

The combination of MFCC and ASS, ASC, ASF descriptors, that give promising results, were used to compute the feature from the acoustic signal. ASF generated vectors with 24 coefficients for each signal frame and scalar value was generated by ASS and ASC.

The final MPEG-7 feature vector composed of 26 (24+1+1) coefficients. The extraction of these descriptors was done in Matlab. The final feature supervector was composed of 39 parameters. These supervectors (MFCC_E+ASS+ASC+ASF = 13+1+1+24) were used to describe the input acoustic signal.

The audio data of the acoustic events used throughout our experiments were recorded in relatively quiet environment.

The sound data were recorded with sampling frequency 48 kHz and with resolution of 16 bits per sample. Then, they were split into the training and the testing set. Recordings were manually labeled using Transcriber. The definition of acoustic event and sound of background was specific for this application. Each sound that consist no acoustic event was considered as a background sound.

The HMM models were trained on the 40 recordings of gunshots, with 40 recordings of breaking glass and 11 minutes of different backgrounds. Testing recordings were used to evaluate the detection system. The character of testing recordings is possible to describe as:

- non-overlapping sounds in recordings[1],
- overlapping sounds in recordings[1, 2],
- overlapping sounds in recordings[2], intensity of added backgrounds was reduced by $-3,5dB$ or $-5dB$,
- overlapping sounds in recordings[2], intensity of added backgrounds was increased by $3,5dB$ or $5dB$.

The total duration of testing recordings was 40 seconds.

IV. Experiments and Results

The performed experiments were focused on the detection of breaking glass and gunshots. HTK software toolkit was used in this task also. For evaluating the proposed system the measure Audio Event Detection Rate (DR) [%] was used. It is defined by the formula:

$$DR[\%] = \frac{\text{Num. of correct recognized models}}{\text{Num. of all reference models}} \times 100. \quad (5)$$

The results are depicted on the Fig. 3. One states HMM models up to 64 PDFs are trained and then evaluated by the testing recordings. For the gun shot detection, the best results (100%) were reached by the overlapping sounds in recordings with suppressed intensity ($-3,5dB$, $-5dB$) of background sounds. This step was done in Audacity software. Several experiments with the increasing intensity ($3,5dB$ and $5dB$) of the background sounds were performed. In these cases the shot detection rates decreased at the level 11,11% for all testing recordings. The Fig. 3 also describes the detection of breaking glass. The best recognition results were obtained for non-overlapping recordings. The second-best results were occurred by overlapping sounds with increased level ($3,5dB$) of background sounds. The glass detection seems to be less sensitive to the change of the intensity of added background in comparison of the shot detection.

V. Conclusion and future work proposal

The conceptual design of surveillance system based on the detection of abnormal sounds such as breaking glass and gun shots was presented in this paper. Presented results give us some basic information about acoustic event detection using MFCC and some of MPEG-7 descriptors. Generally, the high number of PDFs did not bring the system improvement. In the future, we would like to improve the feature extraction methodology e.g. using full set of MPEG-7 descriptors and other parameters like Spectral Flux, Kurtosis, etc. and then we would like to apply PCA feature reduction method. Extension of the Joint Database of Audio Events (JDAE-TUKE) is also in progress.

[1]The reference value of intensity is 0 dB.

[2]Two sounds are hereable in the same time
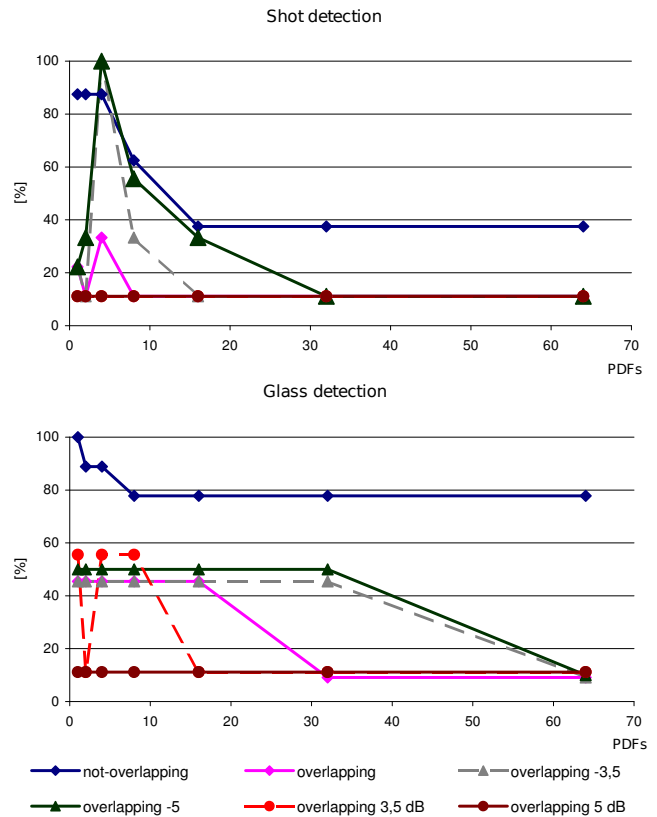
Fig. 3. Detection results.

References

[1] M. Cristiani, et al., Audio-visual event recognition in surveillance video sequences, IEEE Transactions on Multimedia, vol. 9/2, February 2007.
[2] S. Ntalampiras, I. Potamitis, N. Fakotakis, Automatic recognition of urban environmental sounds events, In: CIP proc., pp. 110-113, Santorini, Greece, June 2008.
[3] J.-L. Rouas, J.Louradour, S. Ambellouis, Audio events detection in public transport vehicle, In: ITSC, pp. 733-738, Toronto, Sept. 17-20, 2006, ISBN:1-4244-0093-7.
[4] S. Ntalampiras, I. Potamitis, N. Fakotakis, On acoustic surveillance of hazardous situations, ICASSP, pp. 165-168, Taiwan, April 2009, ISBN: 978-1-4244-2353-8.
[5] M. Ghulam, et al., Environment recognition using selected MPEG-7 audio features and Mel-Frequency Cepstral Coefficients, International Conference on Digital Telecommunications, pp. 11-16, Athens, 2010.
[6] F. Zheng, G. Zhang, Z. Song, Comparison of different implementations of MFCC, Journal of Computer Science and Technology, vol.16/6, November 2001.
[7] E. Vozarikova, A. Cizmar, Speech based features applied to the detection of non-speech audio events, In: RTT, Velk Losiny, pp. 125-128, 2010, ISBN 978-80-248-2261-7.
[8] D. Mitrovic, M. Zeppelzauer, H. Eidenberger, Analysis of the data quality of audio descriptions of environmental sounds, Journal of Digital Information Management, vol. 5/2, pp 48-55, April 2007.
[9] M. Casey, General sound classification and similarity in MPEG-7, Organised Sound, 6 (2), pp. 153-164, 2001.
[10] H.G. Kim, N. Moreau, T. Sikora, MPEG-7 audio and beyond: Audio content indexing and retrieval,Wiley,pp.304,2005, ISBN:0-470-09334-4.
[11] J. Psutka, L. Mller, J. V. Psutka, Comparison of MFCC and PLP parametrizations in the speaker independent continuous speech recognition task, Eurospeech, pp. 1813-1816, Aalborg, September 3-7, 2001.
[12] Young, S., et. al.: "The HTK Book", Cambridge University, 2009.

Drift Correction of the Autonomous Vehicle with the 2D Laser Radar

[1]*František BANÍK,* [2]*Mišél BATMEND*

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]frantisek@banik.sk, [2]misel.batmend@tuke.sk

Abstract—The calculation of the actual position in the vehicle navigation is problematical, because there is a drift error of the wheels in the curve trajectory. The drift correction can be accomplish in a few approaches. This arcticle describes the drift correction based on the second scanning and on the comparison of the two maps scanned from two different positions in the space. The second position is corrected on the parameters determinated from the comparison.

Keywords—Autonomous navigation, Drift correction, 2D map

I. INTRODUCTION

The autonomous vehicle can be used for delivery purposes and for the exploration of the unknown spaces. For the both purpose there is request for the correct definition of the vehicle position in the global coordinate system [1], [2], [3], [4], [5], [6]. The actual position of the vehicle is calculated from wheels, but there is a drift error in curve trajectory. There is a problem for definition the correct vehicle position. Error can be removed by the nonlinear observer, or by the next scanning of the space. The second approach is more robust and correct.

II. VERIFICATION OF THE CALCULATED POSITION

The actual position of the vehicle is defined by the parameters x, y, ψ. Distance moved by the left θ_1 and right θ_2 wheel is the base for calculation of these parameters. The calculation is described as:

$$C_k = C_{k-1} + \left(\frac{\theta_1 + \theta_2}{2} \right) \tag{1}$$

$$\psi_k = \arcsin \left[\frac{\theta_1 - (C_k - C_{k-1})}{\sqrt{[\theta_1 - (C_k - C_{k-1})]^2 + 0,09}} \right] \tag{2}$$

$$x_k = x_{k-1} + [(C_k - C_{k-1}) \cos \psi] \tag{3}$$

$$y_k = y_{k-1} + [(C_k - C_{k-1}) \sin \psi] \tag{4}$$

Where C_k is the shift of the vehicle centre of gravity and the constant $0,09$ is defined by the vehicle width, respectively by the distance between the centers of the wheels. The next position of the vehicle is calculated in discrete process, consequently there is cumulative error in calculation of the position, because the next position is calculated from the previous position $(x_{k-1}, y_{k-1}, \psi_{k-1})$.

Calculated position of the vehicle was verificated in manual move in the space, according to trajectory shown on figure 1 by blue colour. The end calculated position is in point V and the real position is in point S.

The cumulative error is mostly imported in curved trajecotry. The calculated position is relative correct in short shift,

Fig. 1. Comparision of the calculated and real position of the vehicle

consequently in long shift there is the huge difference between calculation and real position. There is the place for correction rules, either developed on the next scanning of the space, or the nonlinear correction in position calculation.

III. THE CORRECTION OF THE POSITION ANALYSIS

The laser radar is the base intecation element of the vehicle with the surrounding space. Every point of the surrounding space is recalculate into the global coordinate system (x, y) after the scanning. The vehicle position and displacement is implemented in recalculation. Recalculation from the polar coordinates, for every point in space, is defined as:

$$x_b = x_k + D_b \sin(\theta_b + \psi_k) \tag{5}$$

$$y_b = y_k + D_b \cos(\theta_b + \psi_k) \tag{6}$$

Where x_k and y_k is the actual calculated position of the vehicle in global coordinate system, based on distance moved by wheels. D_b is the distance between the vehicle and the point on the object. θ_b is the scanner angle for this point on the object. This angle is relative to the vehicle not to the global coordinate system and ψ_k is the actual angle of the vehicle displacement in global coordinate system.

The cumulative error of the position calculation interference into every recalculation of the object positions in the next scans. Two maps are shown on figure 2, the first was scanned from zero coordinates, the second was scanned after vehicle move in the space following axle y. The cumulative error was recalculate into z points of the second map, which is shifted according to the first map, fig 2. Objects in the space were static.
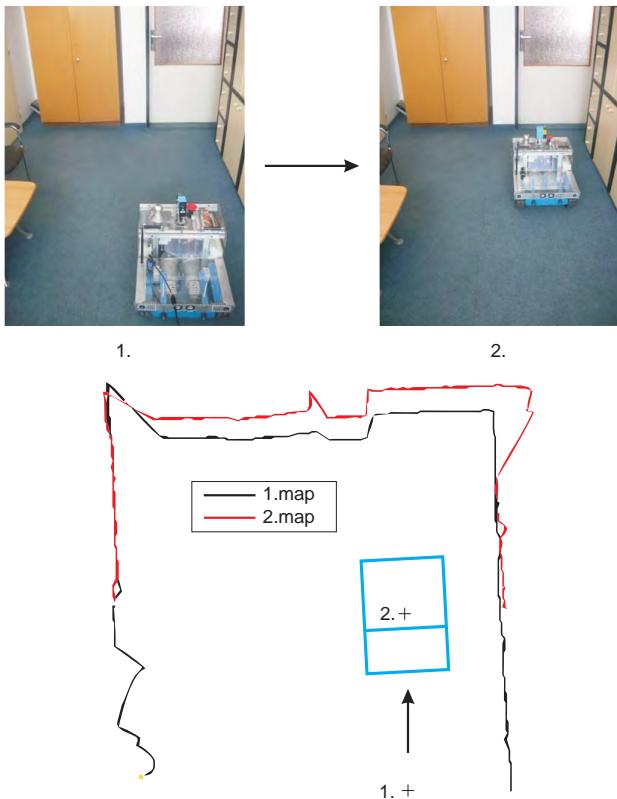
Fig. 2. Interference of the cumulative error into the recalculation of the second map points

Conclusions of this investigation: measurement of the laser scanner is not influence by the cumulative error, precision of the measurement is constant. Cumulative error is influence into recalculation of the points to the global coordinate system. The drift correction can by based on shift between the maps (fig. 2). The other method based on calculation from wheels are not sufficient robustness. Also in the case of the uknown position, there can be find the global position by once more scanning.

IV. PROFILE FUNCTION OF THE MAP

Profile of the map need to be transfered into function desirable for comparison with other map. Profile function represented contours and distance of the map. Also this representation must by independent to the global coordinate system $[x, y]$. For this purpose the angle γ was calculated as is shown on figure 3. The angle γ is from interval $< -\pi; +\pi >$, at which zero represents immediate direction between two lines.
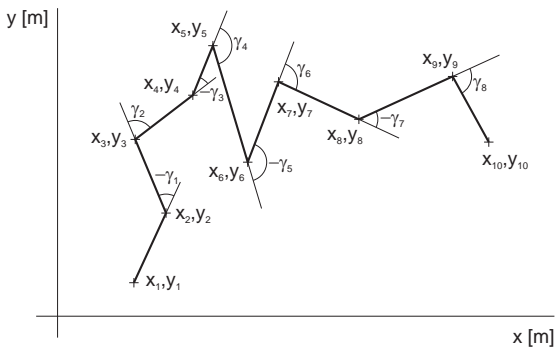
Fig. 3. Angles between lines in map

For the practical purpose was modified the space in office, as is shown on figure 4.

Fig. 4. Modified space in office for demonstration of profile function

The set of points was obtained by scanning of this space, fig. 5 represents the map. Edge $z1 - z9$, dummy edge $f1 - f3$ were mark up in this map. Distances $d1 - d21$ are between this edges.
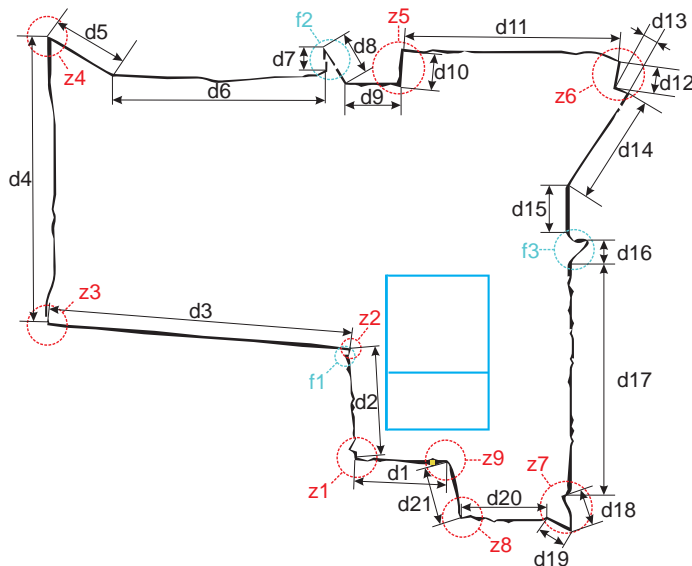
Fig. 5. The line map of the modified space in office, whith marked edges and distances

Demonstration map include 205 lines. The calculated angles between lines are shown in graph 6 a). The edges are marked in this graph in reply to map 5. The calculated angles γ_i were add into summator and for each i is shown in graph 6 b). Also in this graph were marked edges. Conclusions: the start of the summator is zero - start of the map is in zero, the maximal value of the summator is 2π in the case if map is closed. Graph $\sum \gamma_i$ describes the edges in the map, but not describes the distances between them. This fact is the most shown between edges $z2$ and $z3$, where is discontinuity in the map (due to character of the two-dimensional scanner) and between edges is only one line with the distance $d3$.

Correct representation of the map profile must including also the distances between the map edges. Conversion from variously line distance to constant line distance was developed for this purpose. The best constant division for this purpose is $10 \ mm$, that was set by practical testing. This parameter take effect on the last comparision of the map and on the calculation of drift correction. Circuit of the map was divided to the lines with distance $10 \ mm$ and for each line was rewrite $\sum \gamma_i$ due to this section. The result graph is shown on figure 7. Graph
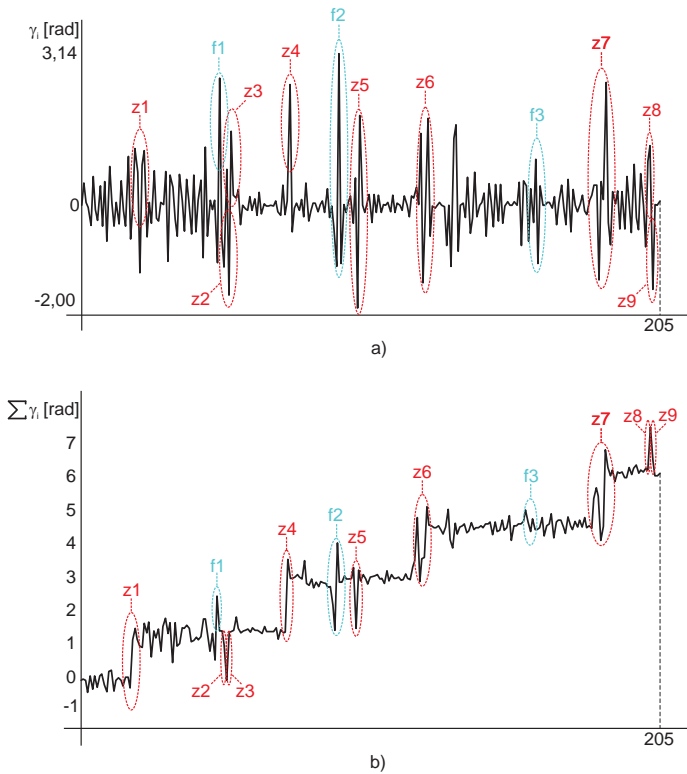
Fig. 6. a) calculated angle γ for each line in the demonstration map, b) summator γ in each line of the demonstration map

of the value τ for the circuit representes profile function of the map, that includes the edges of the map and also the distances between them. The function is independent to the shift and the rotation in the global coordinates system. Profile function is desirable for comparison of the maps.
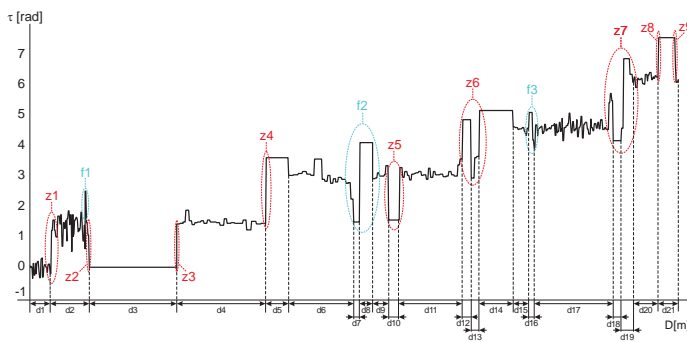
Fig. 7. Profile function τ of the demonstration map

Implementation of the distance $d3$ between the edges $z2$ and $z3$ is shown on fig 7 in confrontation with graph $\sum \gamma_i$ on figure 6 b). Also between edges $z8$ and $z9$. The dummy edges $f1 - f3$, which were created by spaces between wall and furniture and from another point of view need not be created. This edges are represented as oscillate in the result profile function τ, consequently in comparison of the maps this not take the problem effect. This result is important for ideas about another map description, as is constant division, where this dummy edges must be represented as individual point in the map.

V. COMPARISON OF THE TWO MAPS

The simplified space was created to comparison of the two maps 8-1.

Fig. 8. The simplified space for comparison of the map, 1. vehicle position in the center of coordinates, 2. vehicle position after application error of the position

The space was scanned in front of the vehicle in agle from -80 to +80 degrees. The result is vector map shown on figure 9 by black color.
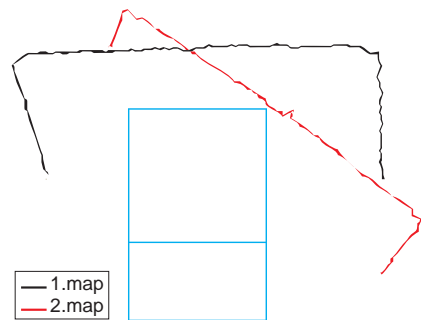
Fig. 9. The result of the two maps mutually moved and swing out for simplified space in global coordinates

In this moment there was inserted the huge error to actual vehicle position. The vehicle was manually moved and twisted to the second position as is shown on figure 8-2. But the actual position in the vehicle controller is not changed and it is in the center of coordinates, as was in scanning of the first position. The second scanning in front of the vehicle was in angle from -30 to 110 degrees. To the calculation of the coordinates for each point in global system is imported the vehicle position, equation 6, and the second map is drawn by red colour. The described method need to change the vehicle position from the center of coordinates to the correct position in the map (shown on fig. 9) into the real position between objects (figure 8-2). Both of the maps were processing as is described in chapter IV. Profile function τ_1 and τ_2 are shown on figure 10.

The comparison of the two maps is to find the point on the first map where can starts the second map, if this second map will be moved and twisted to the position that the lines of the both maps can overlap. The difference value Δ was created to find this point. This value is the surface between graphs τ_1 and τ_2 and is described as:

$$\Delta_k = \sum_{i=1}^{i=i_2} \left| \tau_{1_{i+k}} - \tau_{2_i} - \left(\frac{\tau_{1_k} + \tau_{1_{k+1}} + \tau_{1_{k+2}} + \tau_{1_{k+3}}}{4} \right) \right| \tag{7}$$

Where i_2 is the number of lines of the second map. If $k = 1$ the difference will be calculated for the first case, where the first point of the second map is overlap to the first point of the first map. And the next there are calculation to $k = i_1$ (in our example 227), as is shown on the figure 11.

The moving of the second profile function in τ axle is defined as average of the four values of τ_1, on figure shown by green colour. The average was set by real testing in variously scenarios. If the end of the first map overlap the end of the
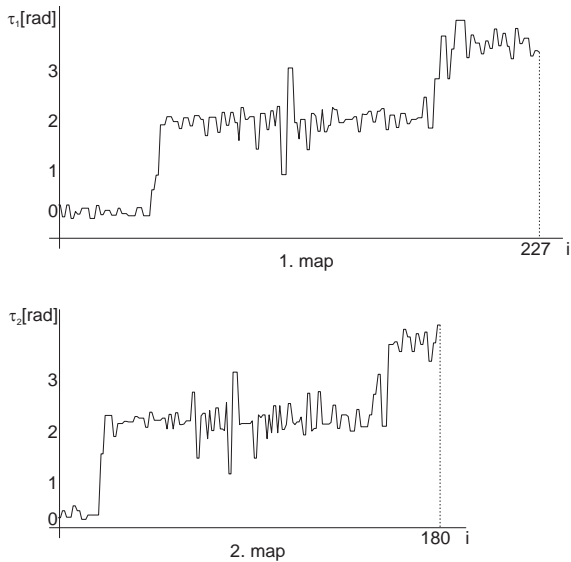
Fig. 10.   Profile function of both maps for comparison
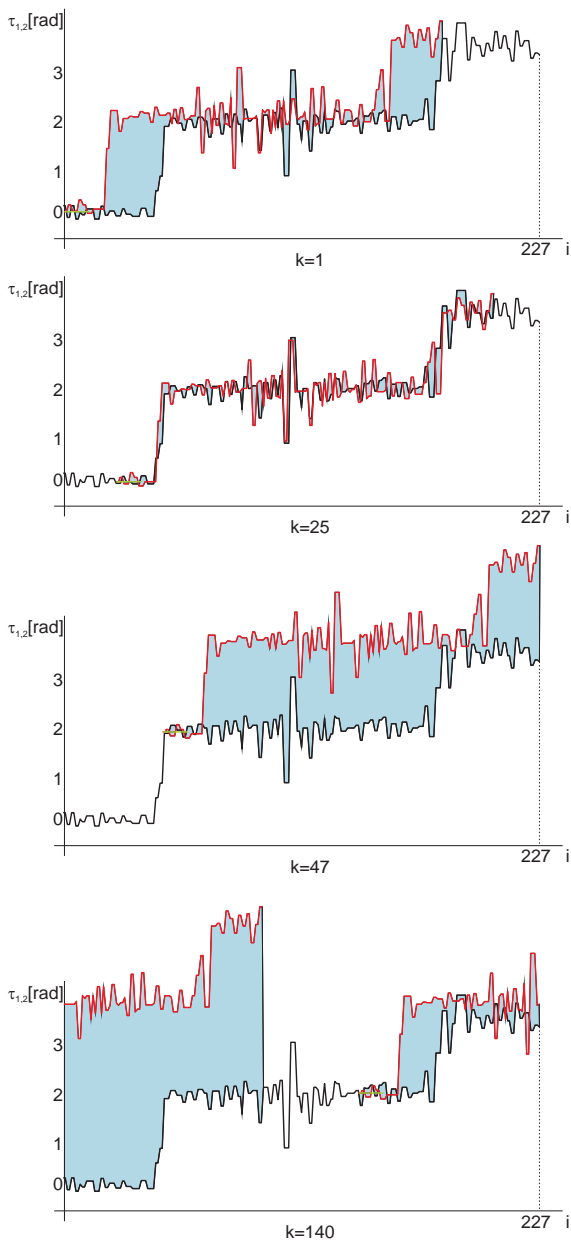
Fig. 11.   Profile functions in calculation of the difference

fisrt map, the calculation continue on the start of the profile funcion τ_1 (fig. 11 for k=140). This attend to passing over the end of the first map. This is utilized only for closed maps, but there is no problem in non closed maps, because in this point cannot be the global minimum of the difference function. Calculation for the all points of τ_1 is shown on figure 12.
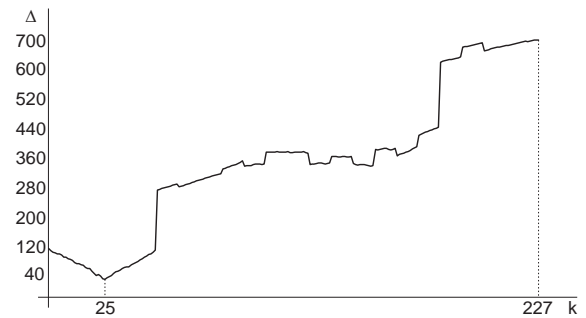
Fig. 12.   The result difference Δ for example maps

The result of the calculation of difference function is, that the minimal difference is in global minimum of the graph 12. The minimum is in point $k = 25$, there is the result from the example graph. That is correct due to compare with figure 11, where for $k = 25$ is the surface between τ_1 and τ_2 the smallest.

The result of these comparison is that the biggest accordance between the maps is if the first point of the second map is set to the point on the first map in distance $25 \times 10mm$ from the start. $10mm$ is the constant of devision.

VI. CONCLUSION

The described method for comparison of two maps brings the numerical result for drift correction of the vehicle position. That was tested on the physical model of the autonomous vehicle [7]. The method is sufficiently robust for laser scanner data. The drift correction is sufficiently correct for the dimensions of the tested vehicle.

REFERENCES

[1] S.-Y. Yi and B.-W. Choi, "Autonomous navigation of indoor mobile robots using a global ultrasonic system," *Robotica*, vol. 22, no. 4, pp. 369–374, 2004.

[2] P. Hoppen, T. Knieriemen, and E. von Puttkamer, "Laser-Radar based Mapping and Navigation for an Autonomous Mobile Robot," *IEEE International Conference on Robotics and Automation 1990*, pp. 948 – 953, May 1990.

[3] S. Clark and H. F. Durrant-Whyte, "Autonomous land vehicle navigation using millimeter wave radar," *International Conference on Robotics and Automation (ICRA)*, pp. 3697–3702, 1998.

[4] S. Clark and G. Dissanayake, "Simultaneous localisation and map building using millimetre wave radar to extract natural features," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1316–1321, May 1999.

[5] H. Hu and D. Gu, "Landmark-based navigation of industrial mobile robot," *International Journal of Industrial Robot*, no. 27, pp. 458–467, 2000.

[6] K. Sugihara, "Some location problems for robot navigation using a single camera," *Journal of Computer Vision, Graphics and Image Processing*, vol. 42, pp. 112–129, 1988.

[7] F. Bank, "Model of the autonomous tracked vehicle," *8th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Kosice*, vol. 4, no. 1, pp. 118–121, May 2008.

Error Concealment in Video Coding using Image/Video Inpainting Method

Branislav HRUŠOVSKÝ, Martin LIPTAJ, Radovan BLICHA

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

branislav.hrusovsky@tuke.sk, martin.liptaj@tuke.sk, radovan.blicha@tuke.sk

Abstract— **Video transmitted over unreliable environment like wireless channel or any network with unreliable transport protocol, is facing the losses of video packets due to network congestion and different kind of noises. To reduce the undesirable effects of information losses, the lost data is usually estimated from the received data, which is generally known as error concealment problem. This paper presents a brief introduction to Image/Video Inpainting methods as a subset of error concealment techniques. Applications of this techniques include the restoration of old photographs and damaged films, the removal of superimposed text like dates or subtitles, the removal of entire objects from the image and other.**

Keywords—**error concealment, image/video inpainting, macroblock.**

I. INTRODUCTION

Motion-compensated hybrid discrete cosine transform (DCT)/differential pulse code modulation (DPCM) coding such as the H.263 and MPEG-1, -2 and -4 are widely used video compression standards. They offer high compression ratio so that video information can be stored or transmitted efficiently. These codecs reduce temporal, spatial and statistical redundancy through motion compensation, quantization of DCT coefficients, and variable length encoding (VLC, for example using Huffman Coding). Comparing with the uncompressed original signals, the compressed digital images and videos are significantly less fault-tolerant to information loss during transmission in noisy channels, and it leads to objectionable visual distortion at the decoder. That type of decoder is possible to design using technology described in [10.] For example, if VLC is used, single bit error may destroy the synchronization of the coded information so that some of the following bits are undecodable until the next synchronization codeword appears. This may cause error propagation over an entire group of pictures (GOP) because of the use of motion prediction/compensation. To reduce this problem, channel coding using unequal error protection (NEP) is frequently used to reduce the adverse effects of channel errors on important information such as motion vectors (MVs) in the coder. To avoid the excessive increase in bandwidth due to channel coding, occasional errors are unavoidable. Therefore, methods for concealing these errors become very important. Basically, the lost data is estimated from the received data. This is generally known as the error concealment problem [6].

II. ERROR CONCEALMENT

Error concealment in block-based decoders has been approached from two primary directions [8]. The first is temporal concealment, which attempts to use the data from past image frames to fill in the corrupted blocks. Simple temporal concealment schemes simply copy the missing macroblocks from a previous frame, while more advanced methods use motion compensation to improve performance [3]. Another technique is spatial concealment, which uses surrounding pixels to reconstruct lost data. In principle, temporal concealment cannot conceal all errors satisfactorily because of occlusion and scene changes. Therefore, an effective spatial concealment algorithm is still required in a complete error concealment system. Spatial concealment based on interpolation works well in flat regions, but usually perform unsatisfactorily at textured regions or edges. Directional interpolation and directional filtering perform much better in these areas. There are also attempts to combine the merits of spatial and temporal methods, which are generally based on mode selection for frames or macroblocks.

A. Combination of Spatial and Temporal methods

As mentioned above, motion of macroblock is critical for temporal error concealment which performs well only in areas of no apparent motion or very consistent motion. This issue, however, does not affect the performance of spatial error concealment. On the other hand, spatial methods are sensitive to content of macroblock and perform well only in smooth areas, while there is no such requirement for temporal methods. So it is desirable to combine the advantages of temporal and spatial methods to achieve a better quality. This is usually achieved by a mode selection mechanism between the two methods as illustrated in Fig.1, for frames or macroblocks. The mode selection using spatial or temporal methods for each video frame is based on scene changes or the frame's type, i.e., I-Frame or P-Frame [5]. In [7], the mode was selected for macroblocks based on the coding modes, i.e., intra-coded or inter-coded. More recently, the mode selection for macroblocks is performed using gradient-based boundary matching (GBM) directional information around the missing macroblock. Unfortunately, these mode selection methods for macroblocks do not always guarantee satisfactory result. For example, if the motion field varies considerably in a

macroblock, we will have to accept either the discontinuity of macroblocks by assigning a unique MV using temporal method or the blurred estimation from spatial method.
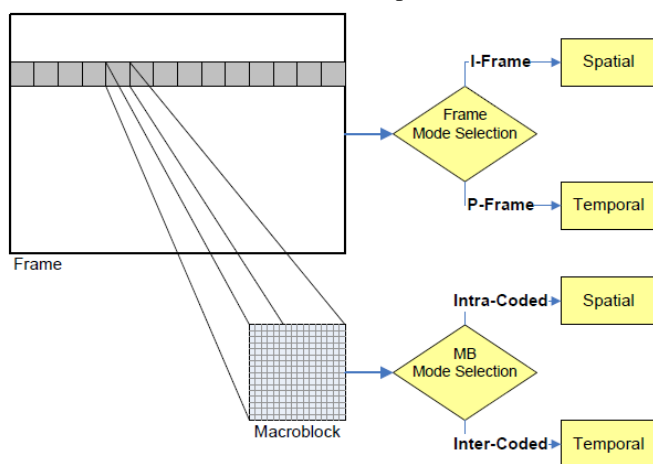
Fig. 1. Combination of temporal and spatial error concealment with mode selection

III. INPAINTING

Interest in the video inpainting field has increased significantly amongst the research community over the past several years due to the varied and important applications of an automatic means of video inpainting. Some of the key applications include:

- *Undesired object removal:* Static or dynamic objects may not be wanted in a film, perhaps due to a change of heart by the director or simply because it was infeasible or unavoidable to exclude the objects during the initial recording (for example, a jet aeroplane flying during the recording of an 18th century drama.)

- *Visual story modification:* Video inpainting can also be extended to change the behaviour of entities within a video. This may be required, for example, to censor an obscene gesture or action that is not deemed appropriate for the target audience, but for which it would be infeasible or expensive to reshoot the scene.

- *Video restoration:* Videos can be damaged by scratches or dust spots or frames could simply be missing or corrupt. It is also possible that during transmission over unreliable networks, information pertaining to significant portions of video frames may be lost in transit. To view the video again in its former glory it is necessary to repair these damaged scenes in a manner that is visually coherent to the viewer [4].

These processes are frequently performed by restoration professionals in a manual fashion, which is not only painstaking and slow but also rather expensive. Therefore any means of automation would certainly be of benefit to both commercial organizations (such as broadcasters and film studios) and private individuals that wish to edit and maintain the quality of their video collection [2].

When an area of an image is missing, inpainting techniques can be used to fill in the gaps. The term "inpainting" comes from art restorers, who often need to fill in parts of a painting that have cracked or flaked off over time. Digital inpainting techniques can be used for all kinds of repairs, such as removing text from an image, erasing powerlines from a scenic view, or repairing cracks and scratches [1].

There are a wide variety of inpainting techniques, all developed in the last few years. One of the simplest is called *"diffusion-based inpainting"*. "Diffusion" is the process by which gas spreads out to fill any volume. For example, when a bottle of perfume is opened, even in the absence of any wind, the smell soon spreads through the whole room, though it is most intense near the bottle. To inpaint, then, we can allow the colors to diffuse into the missing areas of the image [2].

Another way of putting it is that we want to blur the colors out into the missing areas. Mathematically, repeated blurring and diffusion are identical. The technique is known as *"convolution"*. When an image is blurred, the colors of each pixel are averaged with a small portion of the color from neighboring pixels. That pixel, in turn, contributes a small part of its color to each of its neighbors. The image to the left shows color that has diffused from pink and blue spots.

IV. THE MOTION-IMAGE INPAINTING FOR ERROR CONCEALMENT

A. Spatial Methods

Generally, conventional spatial error concealment methods fill the missing MB by interpolating from neighboring pixels around the missing MB, and generate blurred result, so they work well in smooth areas, but not satisfactory for areas containing important structure information such as edges. This problem is also an important topic for image inpainting technology, which is an active research area in recent years. There are some methods proposed in inpainting technology aimed at keeping the structure information, which may be suitable for spatial error concealment [3].

Image inpainting generally refers to spatial image inpainting methods where information from known pixels in the same image is used to inpaint the missing areas more precisely, given an input image I with unknown or missing region Ω, the goal of spatial image inpainting is to propagate information from the known or existing regions $I - \Omega$ of current image to Ω. Various information have been proposed for performing spatial methods. For example, smoothness or Laplacians are used *in partial differential equation (PDE) based method*, and structure or texture are employed for *exemplar-based method*.

In the PDE-based image inpainting method the holes in the image to be inpainted are filled by propagating continuously image Laplacians in the isophote direction from the exterior. The method has its root in the Navier-Stokes equation in fluid dynamics where the known pixels are treated as certain viscous fluid which flows slowly into the missing area Ω as time increases. By discretizing the fictitious time variable by $n.\Delta t$, the image pixel at this time instant can be written as (1):

$$I^{(n+1)}(i,j) = I^{(n)}(i,j) + \Delta t I_t^{(n)}(i,j), \forall (i,j) \in \Omega \qquad (1)$$

where the superscript n denotes the time index n, (i,j) are the pixel coordinates and $I_t^{(n)}(i,j)$ stands for the update of the image $I^{(n)}(i,j)$. Note that the evolution equation runs only inside Ω, the region to be inpainted.

Although PDE-based inpainting method tries to keep direction information by propagating image Laplacians in the isophote direction, it is not applicable for keeping clear structure information. The image inpainting techniques

mentioned above work at the pixel level, and they have worked well for small gaps, thin structures, and text overlays.
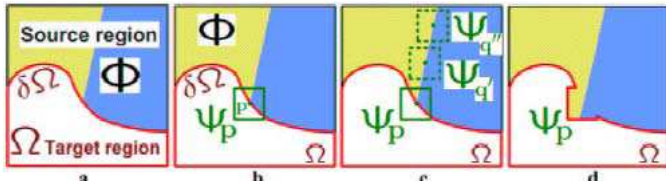
Fig. 2. Structure propagation by exemplar-based texture synthesis

However, for larger missing regions or textured regions, they may generate blurring artifacts. Exemplar-based approaches have also been proposed for image completion by synthesizing pixels using texture synthesis techniques [1]. Recent exemplar-based methods work at the image patch level. They fill in unknown regions more effectively by texture synthesizing with some automatic guidance. This guidance determines the synthesis ordering, which significantly improves the quality of completion by preserving some salient structures. Fig. 2 illustrates the general idea of exemplar-based approaches.

Fig. 2a shows an original image with the target region Ω, its contour $\delta\Omega$ and the source region Φ clearly marked. We want to synthesize the area delimited by the patch Ψ_p centered on the point $p \in \delta\Omega$ (see Fig. 2b). As shown in Fig. 2c, the most likely candidate matches for Ψ_p lie along the boundary between the two textures in the source region, e.g., Ψ_q' and Ψ_q''. The best matching patch in the candidates set is copied into the position occupied by Ψ_p (see Fig.2d), and achieves partial filling of Ω. Finally the target region Ω shrinks and its front assumes a different shape.

It can be seen that spatial methods achieve the objective by preserving or maintaining the image structure or smoothness information in the missing areas. The advantage of spatial methods is its relative simplicity and in the case of image sequences, there is no need to store the extra images since the information is propagated from the current image. On the other hand, since the information in neighboring video frames is not employed, the performance may be significantly affected which may result in images blurring or even incorrect results [4].

When applying exemplar-based method to error concealment, it works well for simple structure such as lines or edges, but still cannot guarantee satisfactory result for complex structure such as cross edges. Also it will fail when there are no similar patches in the frame.

B. Temporal Methods

As mentioned above, conventional temporal error concealment technologies assign one motion vector for all pixels in a frame or macroblock, which is incorrect in most cases. So these techniques perform well only in areas of no apparent motion or very consistent motion but not applicable for complex motion. Fig.3 shows an example of complex motion where traditional temporal error concealment cannot generate satisfactory result.

As shows Fig.3, it is impossible to assign a constant motion to the whole frame because there are different motions for background, shoulder and face. Also it is impossible to assign

a constant motion for MB in the face area because the motion of face is complex.

Fig. 3. Motion field of frame No.193 of video sequence "Foreman"

As shows Fig.3, it is impossible to assign a constant motion to the whole frame because there are different motions for background, shoulder and face. Also it is impossible to assign a constant motion for MB in the face area because the motion of face is complex.

To overcome the shortcoming of traditional temporal error concealment technologies, there must be approaches to assign different motion for each pixel, instead of a constant motion for all pixels in a MB. This can be solved by motion inpainting technology [9].

V. MOTION INPAINTING

Motion inpainting is first introduced as a technique for video stabilization in the computer vision community [2]. By propagating MVs at the outer boundary of missing area into its interior, the motion field in the missing area could be estimated.

As illustrated in Fig.3, the boundary ∂M will gradually advances into the missing area M to fill in the pixel value and motion vectors until it is completely filled [1]. Using known neighboring pixel q_t, whose motion vector is defined by either the initial local motion computation or prior extrapolation of motion data, motion vectors of p_t on the inner boundary of a missing area M are estimated using (2).
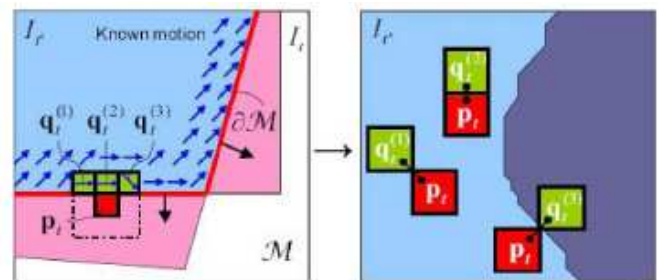
Fig. 4. Motion inpainting

As show Fig.4, motion field is propagated on the advancing front ∂M into M. The color similarities between pt and its neighbors cr are measured in the neighboring frame I_t after warped by local motion of cr' and they are used as weight factors for the motion interpolation [1].

$$F_t^{t-1}(p_t) = \frac{\sum_{q_t \in H(p_t)} w(p_t, q_t) F_t^{t-1}(q_t)}{\sum_{q_t \in H(p_t)} w(p_t, q_t)} \qquad (2)$$

where $H(pt)$ is the neighborhood of pt and F_t^{t-1} is the motion vector from frame t to $t-1$. The motion value for point p_t is a

weighted average of the motion vectors of the points in *H(pt)*. *w(pt, qt)* determines the contribution of the motion value of $qt \in H(pt)$ to pixel p_t, and it is measured by the color similarity or intensity similarity of adjacent image pixels, using (3):

$$w(p_t, q_t) = 1/\{ColorDist\ (p_{t-1}, q_{t-1}) + \varepsilon\}) \tag{3}$$

Motion inpainting offers an efficient approach to estimate the MVs of the missing pixels and hence their pixel values when reliable MVs can be estimated from optical flow. However, its effectiveness also depends on the temporal correlation of the video sequences. A joint spatial-temporal approach is therefore highly desirable, especially in video coding, where abrupt scene changes may be encountered [1].

VI. IMAGE INPAINTING

As shown in Fig.5c, the MRF-based diffusion performed in robust motion inpainting provides blurred estimation for regions where motion inpainting fails. To obtain a better result, image inpainting is performed to refine the MRF based diffused region, see Fig.5e.

The exemplar-based approaches complete the image by synthesizing pixels using texture synthesis techniques. In the proposed method, we apply an exemplar-based method in which works at the image patch level [1]. It fills in unknown regions more effectively by augmenting texture synthesis with some automatic guidance, and get significantly improved quality of completion by preserving some salient structures in the image. Fig.5e shows the result after exemplar-based image inpainting. Comparing with Fig.5c, it can be seen that a sharper and clearer result was provided for region blurred by MRF-based diffusion.

Fig. 5. An example (a) of missing macroblock with wrong motion vector estimation

In Fig.5a, the green arrow shows 10 times enlarged MV, b-e are different results being enlarged by 4 times around the missing MB. Fig.5b illustrates error concealment by motion inpainting, 5c result of MRF-based diffusion in robust motion inpainting, 5d MRF-based diffused region, masked in red in the frame, 5e result of exemplar-based image inpainting on the MRF-based diffused region.

VII. CONCLUSION

The Image/Video(motion) inpainting tries to combine the merits of spatial and temporal methods by mode selection for each pixel. Inpainting methods provide improved performance over conventional methods for the videos tested and they serve as an alternative to existing methods for video error concealment. Although the Image/Video inpainting tries to eliminate the influence of inaccurate motion by mode selection

based on robust criteria, it still suffers from motion estimation errors. The challenge lies on robust motion estimation for error concealment, and a more reliable mode selection criterion.

REFERENCES

[1] CH. Liyong, "Joint Imge/Video Inpainting for Error Concealment in Video Coding," M. Phil. Thesis, The University of Hong Kong, December 2007.

[2] M. Bertalmio, G. Sapiro, V. Caselles, C. Ballester, "Image Inpainting", *Proceedings of SIGGRAPH 2000*, New Orleans, USA, July 2000.

[3] K. Raghavendra, "Video Inpainting and Object Separation" Thesis, Department of Electrical Engineering, Indian of Institute of Technology, Kanpur, December 2007.

[4] M. M. Oliveira, B. Bowen, R. McKenna, Y. Chang, "Fast Digital Image Inpainting", *Proceedings of the International Conference on Visualization, Imaging and Image Processing (VIIP 2001)*, Marbella, Spain, September 3-5, 2001

[5] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 560-576, 2003.

[6] S. Shirani, F. Kossentini, R. Ward, "A Concealment Method for Video Communications in an Error-prone Environment," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1122-1128, 2000.

[7] S. Belfiore, M. Grangetto, E. Magli, G. Olmo, "Spatio-Temporal Video Error Concealment With Perceptually Optimized Mode Selection", *ICASSP 2003 - IEEE International Conference on Acoustics, Speech, and Signal Processing*, Hong Kong, April 2003.

[8] K. Jung, J. Chang, C. Lee, "Error Concealment Technique Using Projection Data for Block-based Image Coding," *Proc. SPIE Conf Visual Communic. and Image Proc*, vol. 2308, pp.1466-1476, 1994.

[9] http://iat.ubalt.edu/summers/math/inpainting.htm

[10] M. Liptaj, "Návrh ASIC obvodov pre širokopásmové aplikácie", *Minimal thesis*, Technical university of Kosice, january 2011

Experiments with Principal Component Analysis Applied to Speech Multi-frames

Peter VISZLAY

Department of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

peter.viszlay@tuke.sk

Abstract—**In this paper a linear transformation based feature extraction used in speech recognition is presented. It is applied to speech multi-frames created from basic cepstral vectors with various lengths. The influence of these lengths to recognition rate is investigated. Also the performance of the transformation in dimension reduction is studied. To reduce the dimensionality of the concatenated basic vectors, class-independent Principal Component Analysis (PCA) with reduced rank transformation matrix was used. We have focused on the possible improvement of the recognition performance with the proposed methodology in HMM-based context independent phoneme recognition task. Several experiments using TIMIT speech database and its manual phonetic transcriptions were done. Experimental results for different model dimensions for baseline and transformed acoustic models are evaluated and compared with detailed conclusion and analysis of results.**

Keywords—**dimension, feature vector, linear transformation, matrix, supervector**

I. INTRODUCTION

Feature extraction is an important part of the whole recognition process of any automatic speech recognition (ASR) system, because it greatly affects the recognition performance of the system [1]. The feature extractor has to derive a relevant speech representation resulting in form of D-dimensional feature vectors, which are obtained by applying Fourier or cepstral analysis to short time segments [2]. The conventional acoustic front-ends designed for most of ASR systems are based on MFCC (Mel-Frequency Cepstral Coefficients) or PLP (Perceptual Linear Prediction) features. These techniques are good choices to obtain satisfactory speech representations.

In most ASR systems, the basic feature extraction process is supplemented by linear transformation in order to improve the recognition performance. The feature extraction process is then achieved in two steps: parameter extraction and feature transformation. In general, the linear transformations (LTs) project the mentioned D-dimensional basic feature vectors to d-dimensional transformed subspace ($d < D$) according to some criterion (retaining maximum variance, class discrimination, etc.). LTs reduce the feature dimensionality, so the computational cost and system complexity for subsequent processing is decreased. Linear Discriminant Analysis (LDA) [3] and Principal Component Analysis (PCA) [1] are two popular linear transformation methods.

In this paper, we have built upon our previous work and we present a methodology of application of a data-driven unsupervised feature transformation based on Principal Component Analysis applied to supervectors (multi-frames) created by concatenating of successive MFCC vectors. The experimental results of this method are compared to baseline system in the TIMIT phoneme recognition task. The reference baseline system was built on MFCC vectors that consisted of first few static cepstral coefficients, 0th coefficient and also of the first and second-order derivatives. The specific vector compositions are listed in Table I. In the next section the related work description and in the third section the PCA is presented. Sections IV, V and VI describe the experimental setup, the results and conclusions.

II. RELATED WORK

In this work, we have tried similar approaches that were investigated in other applications under analogous conditions. The TIMIT speech database [4] used in our experiments is an internationally renowned corpus specifically designed to support the development and evaluation of models. Some works [2], [5], [6], [7] and [8] are focused to applications evaluated especially on this database using some additional methods. In [7] the researchers have applied a very similar method (LDA, PCA and LP transformation - concatenating the LDA and PCA coefficients) to transform the feature vectors. We have investigated a similar methodology (only the PCA has been used) with completion of the methodology with concatenating the base vectors used as input for PCA. This similar approach resulted in comparable, in some cases better results than were publicated in the mentioned publications. The supervectors were also used in [6] and they are one of the fundamental operations in LDA application [3] in speech recognition. In [6] the PCA was applied to the multi-frame context windows and it was also applied in acoustic events detection system in feature extraction [9].

III. LINEAR FEATURE TRANSFORMATIONS AND DIMENSIONALITY REDUCTION

Linear dimensionality reduction

Linear feature transformations are used in ASR to convert and reduce the original features to an alternative and more compact set retaining information as much as possible [7]. This is achieved by retaining only the relevant dimensions according to choosed criterion. This step helps to solve the problem called *the curse of dimensionality*. Reducing the dimensionality of features is the most direct way to solve the problems caused by high dimensionalities. A general mathematical model of an LT can be written as:

$$y = W^T x, \qquad (1)$$

where y is the output transformed feature set, W is the transformation matrix and x is the input set.

Principal Component Analysis

Principal Component Analysis is a popular data processing and dimension reduction method applied also in speech recognition. PCA maps the n-dimensional input data to m-dimensional, $m < n$, with respect to the variability of data. The method is based on the assumption that most information about classes is contained in the directions, along which the variations are the largest [7]. But, there is no guarantee that variability explained by PCA is useful for speech recognition [1] (PCA is strongly data-dependent and not all of the data sets are suitable for PCA). It transforms the data by principal components PCs (uncorrelated and ordered variables). Usually, with the first few PCs can be represented about 80% variability of the basic vectors. PCA was used in our experiments in the following mathematical way according to [10] and [11]. Suppose that we have M feature vectors $x_1, x_2, \ldots, x_M$ corresponding to speech signals in the training set. Firstly, the input data have to be subtracted:

$$\Phi_i = x_i - \bar{x} = x_i - \frac{1}{M}\sum_{i=1}^{M} x_i, \qquad (2)$$

where Φ_i is the i-th centered vector and $\bar{x}$ is the mean vector. From these vectors are then created the centered data matrix $A = [\Phi_1\Phi_2\ldots\Phi_M]$. Principal components can be given by K leading eigenvectors of the global covariance matrix C resulting from its eigendecomposition:

$$Cu_i = \lambda_i u_i, \qquad i \in 1, \ldots, N, \qquad (3)$$

where $u_1, u_2, \ldots, u_N$ are the eigenvectors and $\lambda_i, \ i \in \langle 1; N \rangle$ are the eigenvalues of the covariance matrix:

$$C = \frac{1}{M-1}\sum_{n=1}^{M}\Phi_n\Phi_n^T = \frac{1}{M-1}\sum_{n=1}^{M}(x_i - \bar{x})(x_i - \bar{x})^T. \quad (4)$$

The dimensionality reduction step is performed by keeping only the eigenvectors corresponding to the K largest eigenvalues ($K < N$) and put them into matrix $U_K = [u_1 u_2 \ldots u_K]$, where $\lambda_1 > \lambda_2 > \ldots > \lambda_N$. Finally, the linear transformation $\mathbf{R}_N \rightarrow \mathbf{R}_K$ is computed as:

$$y_i = U_K^T x_i, \qquad (5)$$

where y_i represents the transformed vector and U_K is the reduced PCA matrix. The determination of K can be done via comparative criterion with threshold $T \in \langle 0.9; 0.95 \rangle$:

$$\frac{\sum_{i=1}^{K}\lambda_i}{\sum_{i=1}^{N}\lambda_i} > T. \qquad (6)$$

IV. METHODS AND EXPERIMENTS

Control experiment

In order to compare the results of the proposed methodology with the baseline acoustic model it was necessary to build several base MFCC models. The accuracies of this model are listed in Table IV. The feature extraction for MFCC was performed in common way. The input speech signal is preemphasized and windowed using Hamming window. The window size was 25 ms. Fast Fourier transform was applied to the windowed segments. In the next step, the mel-filterbank analysis with 20 channels was applied followed by logarithm application to the linear filter outputs. In order to decorrelate the features, the discrete cosine transform (DCT) was applied.

This acoustic analysis for different dimensions was performed, according to Table I, where the symbol N has varied as follows: N=4, 7, 9, 10, 11, 12, 13, 14 and 15. These vectors were then used as the input for the following transformations.

TABLE I
MFCC VECTORS COMPOSITIONS FOR ALL DIMENSIONS

MFCC vectors	Final dimension	Dimension for PCA
N static + 0th + Δ + $\Delta\Delta$	$n=3(N+1)$	$n_{eff}=N+1$

Proposed methodology

The goal of the described methodology was the application of PCA to multi-frame feature vectors. MFCC feature vectors were used to create supervectors. To learn the PCA transformation matrix only the train set was used. Each speech signal in the train corpus is represented after the parametrization process by data matrix with size $n \times n_i$, where n is the actual dimension (according to Table I) and n_i is the number of frames in i-th recording. By pooling these matrices, the training set can be represented by one big data matrix of size $n \times L$, where $L = \sum_{i=1}^{N} n_i$. The TIMIT training set has N=4620 recordings and L=1 410 069 (this number is the sum of all columns in the sub data matrices). The MFCC multi-frames were used to add context, with appropriate number of frames. The actual feature vector $x[n]$ on position n is always modeled with its left context vectors $x[n-1], \ldots, x[n-(C-1)/2]$ and its right context vectors $x[n+1], \ldots, x[n+(C-1)/2]$, where C is the multi-frame length. For example, when the multi-frame size was set to C=5, the supervector, which models the context for actual vector $x[n]$ has a form $x[n-2], x[n-1], x[n], x[n+1], x[n+2]$.

In our experiments, we have used two different multi-frame lengths; C=3 and C=5. We operated overall with 18 various types of full rank PCA transformation matrices, which had sizes $n_{eff}C \times n_{eff}C$. So, for instance for C=5 and dimension n_{eff}=18 we had PCA matrix with size $U_{[80\times80]}$. At the transformation, their reduced rank versions were used. The specific supervector compositions before computing PCA are listed in Table II. The final dimensions were chosen independently from the criterion (6) in order to compare the performances at the same dimension.

TABLE II
TRANSFORMED PCA VECTORS COMPOSITIONS

PCA vectors, C=3	$3(N+1)$MFCC $\rightarrow$ $(N+1)$PCA + Δ + $\Delta\Delta$
PCA vectors, C=5	$5(N+1)$MFCC $\rightarrow$ $(N+1)$PCA + Δ + $\Delta\Delta$
Final dimension	$n = 3(N+1)$

Adding context to actual vector, when it is inside a data matrix is not problematic. The problem will arise, when its actual position is at the beginning or at the end of the matrix. At the beginning does not exist the left context and at the end does not exit the right context yet. To solve this problem we have done some modifications at these positions. This problem is referred as *start* and *end-effect problem*. We have tried three different approaches to generate the non existing context. First one consisted in zero vectors addition to first and last feature vector in matrix. The second one is some kind of continuation of speech signal. Exactly, the left context vector was created with the same dimension from the first

coefficient of first feature vector and the right context vector was created from the last coefficient of the last feature vector. This approach is interpreted by Fig. 1, where the matrices $X(1)$, $X(2), \ldots,$ $X(N)$ are the basic data matrices. The third

$$\begin{bmatrix} x_{11}^{(1)} \\ x_{11}^{(1)} \\ \vdots \\ x_{11}^{(1)} \end{bmatrix} \ldots \begin{bmatrix} x_{11}^{(1)} \\ x_{11}^{(1)} \\ \vdots \\ x_{11}^{(1)} \end{bmatrix} \begin{bmatrix} x_{11}^{(1)} & x_{12}^{(1)} & \cdots & x_{1n_1}^{(1)} \\ x_{21}^{(1)} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{n1}^{(1)} & \cdots & \cdots & x_{nn_1}^{(1)} \end{bmatrix} \begin{bmatrix} x_{11}^{(2)} & x_{12}^{(2)} & \cdots & x_{1n_2}^{(2)} \\ x_{21}^{(2)} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{n1}^{(2)} & \cdots & \cdots & x_{nn_2}^{(2)} \end{bmatrix} \ldots$$
$$\cdots \begin{bmatrix} x_{11}^{(N)} & x_{12}^{(N)} & \cdots & x_{1n_N}^{(N)} \\ x_{21}^{(N)} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{n1}^{(N)} & \cdots & \cdots & x_{nn_N}^{(N)} \end{bmatrix} \begin{bmatrix} x_{nn_N}^{(N)} \\ x_{nn_N}^{(N)} \\ \vdots \\ x_{nn_N}^{(N)} \end{bmatrix} \cdots \begin{bmatrix} x_{nn_N}^{(N)} \\ x_{nn_N}^{(N)} \\ \vdots \\ x_{nn_N}^{(N)} \end{bmatrix}$$
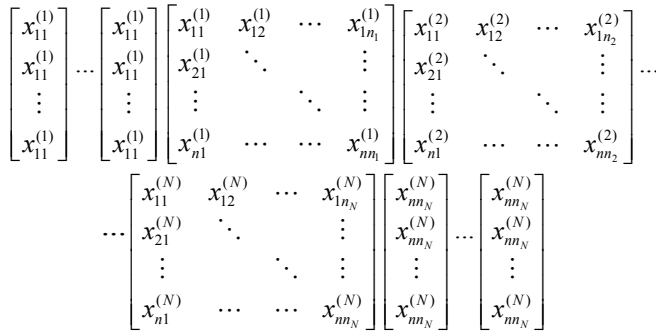
Fig. 1. Illustration of the context vectors adding

approach was simply repeating of the first and last feature vector. After experimental verification of these approaches we chosed the second one, because it provided satisfactory results (performance differences between these three approaches were 0.5-1%).

In the phase of constructing the supervectors their length has varied according to the context size. When the context was set to C=3, then the supervector length was three times longer than the basic vector (analogy with C=5). For C=3 and for actual vector one left context vector and one right context vector was taken to create the corresponding supervector. This process is modeled by Fig. 2 and it resulted in a matrix S (we called it as *supermatrix*) consisting of supervectors in its columns. This supermatrix can be viewed as an analogy with

$$S = \begin{bmatrix} \begin{bmatrix} x_{11}^{(1)} \\ \vdots \\ x_{11}^{(1)} \end{bmatrix} & \begin{bmatrix} x_{11}^{(1)} \\ \vdots \\ x_{n1}^{(1)} \end{bmatrix} & \cdots & \begin{bmatrix} x_{1(n_1-1)}^{(1)} \\ \vdots \\ x_{n(n_1-1)}^{(1)} \end{bmatrix} & \begin{bmatrix} x_{1n_1}^{(1)} \\ \vdots \\ x_{nn_1}^{(1)} \end{bmatrix} & \cdots & \begin{bmatrix} x_{1(n_N-2)}^{(N)} \\ \vdots \\ x_{n(n_N-2)}^{(N)} \end{bmatrix} & \begin{bmatrix} x_{1(n_N-1)}^{(N)} \\ \vdots \\ x_{n(n_N-1)}^{(N)} \end{bmatrix} \\ \begin{bmatrix} x_{11}^{(1)} \\ \vdots \\ x_{n1}^{(1)} \end{bmatrix} & \begin{bmatrix} x_{12}^{(1)} \\ \vdots \\ x_{n2}^{(1)} \end{bmatrix} & \cdots & \begin{bmatrix} x_{1n_1}^{(1)} \\ \vdots \\ x_{nn_1}^{(1)} \end{bmatrix} & \begin{bmatrix} x_{11}^{(2)} \\ \vdots \\ x_{n1}^{(2)} \end{bmatrix} & \cdots & \begin{bmatrix} x_{1(n_N-1)}^{(N)} \\ \vdots \\ x_{n(n_N-1)}^{(N)} \end{bmatrix} & \begin{bmatrix} x_{1n_N}^{(N)} \\ \vdots \\ x_{nn_N}^{(N)} \end{bmatrix} \\ \begin{bmatrix} x_{12}^{(1)} \\ \vdots \\ x_{n2}^{(1)} \end{bmatrix} & \begin{bmatrix} x_{13}^{(1)} \\ \vdots \\ x_{n3}^{(1)} \end{bmatrix} & \cdots & \begin{bmatrix} x_{11}^{(2)} \\ \vdots \\ x_{n1}^{(2)} \end{bmatrix} & \begin{bmatrix} x_{12}^{(2)} \\ \vdots \\ x_{n2}^{(2)} \end{bmatrix} & \cdots & \begin{bmatrix} x_{1n_N}^{(N)} \\ \vdots \\ x_{nn_N}^{(N)} \end{bmatrix} & \begin{bmatrix} x_{1n_N}^{(N)} \\ \vdots \\ x_{nn_N}^{(N)} \end{bmatrix} \end{bmatrix}$$

Fig. 2. Supermatrix composed by splicing the basic vectors

so-called *Hankel matrix*, but in block matrix form. This fact helped us to reduce the processing time. It was then considered as the input matrix for PCA (matrix A labeled in Section III). Then the main PCA was performed according to mathematical description in Section III.

The big matrix A was analyzed by PCA, then the global covariance matrix was computed and finally, its spectral decomposition was done, which resulted in a set of eigenvectors and eigenvalues. The number of eigenvectors was identical with the number of input dimension for PCA ($n_{eff}C$). From the eigenvectors were chosen the leading ones corresponding to the largest eigenvalues. The eigenvalues rapidly decrease with their index so the first few eigenvalue - eigenvector pairs are sufficient to retain the most variance present in the data. The dominant eigenvectors then formed the reduced

rank transformation PCA matrices. With these matrices the train and test sets were transformed independently. The actual PCA matrix was learned in the PCA training process (Fig. 3). In order to visualize the matrix coefficients in magnitude
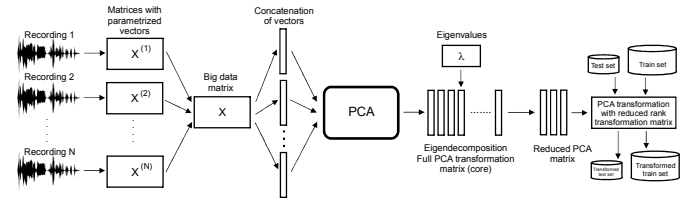
Fig. 3. Learning process of full rank PCA matrix

sense it was randomly selected one recording from test corpus. Fig. 4 on the left shows the matrix for 10 MFCCs with approximately 400 feature vectors. The maximum magnitude is roughly 20 (white areas). On the right side the same matrix is depicted after PCA (C=3) with the same dimension. There is concentrated more energy (lighter areas; maximum magnitude varies around value 40 - this is twice more than previously). All processing related to PCA in the MATLAB environment
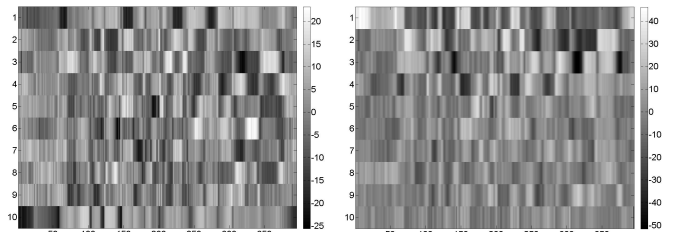
Fig. 4. MFCC and transformed coefficients

was performed. Using special functions from VOICEBOX - A Speech processing Toolbox for MATLAB [12] we achieved compatibility between the feature vectors format and MAT-LAB. We have done a few experiments for higher context lengths (C=7, 9 and 11), but in some cases there occurred out of memory problems (a need to allocate really big matrices, with millions of real valued coefficients). For example, in our experiments we have done about 60 training procedures and about 500 different tests and evaluations, but only a half of them were useful for our intentions. These operations take approximately 1050 hours of absolute processing time.

The speech corpus

As was mentioned before, the TIMIT speech database we have used to investigate the proposed methodology. This database is divided into train and test set, with 4620 and 1680 recordings, respectively. There are 61 distinct phones, which were reduced to an inventory of 40 symbols. It was used a phoneme mapping structure according to Table III. All closure phonemes ('bcl', 'dcl', 'gcl', 'kcl', 'pcl', 'tcl') were merged in the training phase with previous phoneme segment.

TABLE III
PHONEME MAPPING SCHEME

TIMIT	ao	ax	axh	axr	hv	ix	el	em	en
Mapped	aa	ah	ah	er	hh	ih	l	m	n
TIMIT	nx	eng	zh	*cl	h#	pau	epi	ux	q
Mapped	n	ng	sh	-	sil	sil	sil	uw	sp

Experimental setup

The HMM parameters for 40 phonemes were trained. Each phoneme was modeled by a three-state left-to-right HMM. Probability density functions (PDF) were used to model the HMM states with 1, 2, 4, 8, 16, 32, 64, 128 and 256 Gaussian mixtures. This training scheme was applied to each transformation. In order to test the acoustic models a simple language model was built. It is a backoff bigram model smoothed using Witten-Bell discounts. It was composed from the whole database using SRI Language Modeling Toolkit. The feature extraction, acoustic modeling and testing by HTK (Hidden Markov Model) Toolkit were carried out.

V. EXPERIMENTAL RESULTS

All experimental results of our experiments are listed in Table IV (the accuracy for MFCC and PCA model). These results show that the proposed methodology outperforms the MFCC approach. The PCA provides for $C=3$ better recognition performances than MFCC for lower dimensions and for cases with more Gaussian mixtures (64 - 256). On the other hand, recognition performance can be improved again for higher model dimensions, but for lower Gaussian mixtures (improving with gray color is highlighted). The achieved improvement of accuracy is approximately 1%. For the case $C=5$ the obtained results are less satisfactory. It can be seen that for longer contexts we have achieved not significant improvement in comparison with shorter contexts. The gray areas in the table mean improvement in comparison with the baseline model and the bold values mean improvement compared to case with shorter contexts. Notice that the accuracies in the evaluation process were computed as the ratio of number of all word matches to number of reference words. It should be noted that the PCA analysis process in MATLAB is memory intensive, but the recognition time of the final recognizer with reduced dimensions is shorter and the memory requirements are lower. This fact may lead to more effective function of the recognizer in real time applications.

VI. CONCLUSIONS AND FUTURE WORK

The main conclusion of our presented work is that the PCA applied to supervectors can outperform the standard MFCC approach for lower dimensions, so it is more effective in energy cumulation for low-order coefficients composing a lower dimensional feature vectors. The second conclusion is that for longer contexts the performances are not improved. This fact we can interpret it so that the longer multi-frames are convenient for larger speech databases than TIMIT.

In the future, we want to refine the PCA algorithm (other mathematical methods used in covariance matrix decomposition) and we want to explore the context adding to start and end of the data matrix. We also want to implement the PCA on a larger database with real Slovak speech data.

ACKNOWLEDGMENT

TABLE IV
ACCURACIES FOR MFCC AND PCA-BASED MODELS

Dim.	15	24	30	33	36	39	42	45	48
Mixtures	*Accuracy (%) for baseline MFCC model*								
1	62.38	66.57	67.96	68.01	67.89	67.54	67.43	67.09	66.94
2	64.18	67.42	69.02	68.86	68.81	68.85	68.61	68.57	68.43
4	66.26	69.22	70.79	71.35	70.78	71.17	70.90	70.89	70.75
8	67.27	70.63	72.26	72.54	72.92	72.69	72.54	73.10	73.66
16	67.92	71.94	73.36	73.57	73.98	74.14	74.20	74.69	74.96
32	68.87	72.49	74.37	74.71	75.69	75.43	75.76	75.80	76.48
64	69.52	73.32	75.01	75.83	76.47	76.43	76.71	76.98	77.22
128	69.69	73.58	75.44	76.32	76.70	77.05	77.36	77.50	77.79
256	70.53	74.41	75.96	76.28	76.85	77.27	77.43	77.16	
	Accuracy (%) for PCA with supervectors from multi-frames, C=3								
1	62.29	65.49	67.38	67.46	67.53	67.33	67.19	67.06	67.10
2	63.82	67.05	69.44	68.67	69.30	69.80	68.74	69.20	69.02
4	65.26	68.33	71.19	70.33	71.26	71.03	71.02	71.31	71.72
8	66.79	69.66	72.06	72.25	72.68	72.84	73.29	73.07	73.23
16	67.64	70.88	73.54	73.52	73.86	74.29	74.63	74.60	74.34
32	68.76	72.39	74.34	74.54	75.16	75.08	75.71	75.34	75.80
64	69.66	73.24	75.05	75.24	75.70	75.96	76.69	76.51	77.09
128	69.99	73.93	75.56	75.64	76.44	76.56	76.85	77.05	77.14
256	70.62	73.99	75.44	75.67	76.28	76.80	77.00	76.72	76.96
	Accuracy (%) for PCA with supervectors from multi-frames, C=5								
1	61.01	64.25	66.26	66.17	66.34	66.04	66.60	66.11	66.77
2	62.09	65.86	67.46	67.84	67.96	68.34	68.11	68.39	68.32
4	64.02	67.67	68.46	69.76	69.83	70.37	70.55	70.54	70.79
8	65.71	69.63	71.36	71.54	71.95	72.41	72.91	72.98	**73.30**
16	66.77	**71.00**	72.77	73.14	**74.00**	73.92	74.48	74.51	**74.85**
32	68.22	**72.48**	73.72	74.31	74.62	75.06	75.49	**75.92**	**76.00**
64	69.29	**73.51**	74.80	75.12	75.35	75.77	76.39	**76.63**	**76.96**
128	**70.33**	**74.02**	75.48	75.61	76.11	76.49	**76.86**	**77.32**	77.63
256	**70.68**	**74.36**	**75.85**	**75.94**	76.11	76.36	76.56	**77.09**	**77.67**

REFERENCES

[1] H. Abbasian and B. A. Nasersharif, "Class-Dependent PCA Optimization Using Genetic Programming for Robust MFCC Extraction," *In: 3rd Conference on Information and Knowledge technology (IKT)*, Mashhad, Iran, 2007.

[2] A. Errity, J. McKenna, and B. Kirkpatrick, "Manifold Learning-Based Feature Transformation for Phone Classification," *LNCS*, vol. 4885, Springer, 2007.

[3] R. Haeb-Umbach and H. Ney, "Linear discriminant analysis for improved large vocabulary continuous speech recognition," *In Proceedings of IEEE ICASSP*, pp. 13–16, San Francisco, CA, March 1992.

[4] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, and N. Dahlgren, "The DARPA TIMIT acoustic-phonetic continuous speech corpus cd-rom," *Technical report*, 1993.

[5] F. Triefenbach, A. Jalalvand, B. Schrauwen, and J.-P. Martens, "Phoneme Recognition with Large Hierarchical Reservoirs," *In Proc. of Neural Information Processing Systems)*, Dec., Vancouver, 2010.

[6] P. Somervuo, "Experiments With Linear And Nonlinear Feature Transformations In HMM Based Phone Recognition," *In Proceedings of ICASSP*, pp. 52–55, 2003.

[7] X. Wang and D. O'Shaughnessy, "Improving the efficiency of automatic speech recognition by feature transformation and dimensionality reduction," *In Proc. of Eurospeech*, pp. 1025–1028, Geneva, Sep. 2003.

[8] M. Katrák, "Klasifikácia foném reči neurónovou sieťou," *PhD. thesis*, Technical university of Košice, Slovakia, 2010.

[9] E. Vozáriková, P. Viszlay, J. Juhár, and A. Čižmár, "Acoustic events detection via PCA-based feature extraction," *In: Journal of Computer Science and Control Systems*, vol. 3, pp. 99–102, No.2, 2010.

[10] G. Bebis, "Principal Component Analysis," online: www.cse.unr.edu/ bebis/MathMethods/PCA/lecture.pdf.

[11] I. T. Jolliffe, *Principal Component Analysis*. Berlin, Germany: Springer Series in Statistics, ISBN: 0-387-95442-2, 1986.

[12] "VOICEBOX: Speech Processing Toolbox for MATLAB," online: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html.

Impact of the Parameters on the Resulting Graph cut Segmentation

Peter KARCH, Marián BAKOŠ, Oľga Duľová

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

peter.karch@tuke.sk, marian.bakos@tuke.sk, olga.dulova@tuke.sk

Abstract — **This article presents the impact of the parameters used in the Graph cut method on the final segmentation. The article shows how the coefficients are implemented in the regional and border properties and their impact on the final segmentation in Graph cut segmentation. Subsequently, it describes the principle of using regional and border properties in this method of segmentation. For demonstration of the impact of coefficients in the parameters, illustrative experimental verification of this method is used.**

Keywords—**Graph cut, parameters, segmentation, maximal flow, terminals, initialization.**

I. INTRODUCTION

One of the basic methods of image processing is image segmentation. These segmentations are required to separate objects in the image from their background. This activity can be performed with the possibly smallest number of user interventions Graph cut segmentation, which belongs to advanced segmentation techniques, needs properly configured parameters for its operation. For initialization, it only needs to identify a few pixels belonging to the objects and a few pixels belonging to the background. Precision can be achieved by an appropriate number of initialization pixels of the object and background segmentation. Another important parameter for its precision is appropriate determination of internal parameters for computation of the regional and border properties for the Graph cut method.

The details of Graph cut segmentation technique are described in Section II. Section III experimentally demonstrated the impact of parameters on resulting segmentations. In Section IV, the impact of these parameters on resulting Graph cut segmentations is evaluated.

II. SEGMENTATION TECHNIQUE

A. Graph cut segmentation

This method divides an image into two segments: "object" and "background", and belongs to the interactive segmentation technique. Segmentations [1], [3] are determined as labeling vector $A = \{A_1,...,A_p,...,A_{|P|}\}$. Segmentation of the objects from its background is defined as binary labeling problem, where every pixel is labeled as $A_k = \{O, B\}$, where O represents the pixel belonging to the object in the image and B represents the pixel belonging to the background

of the image. These binary labels are represented as hard constraints of the segmentation and are identified by the terminal nodes S (source) and T (sink), which represent "object" and "background" of the segmentation. Other pixels in the image represent soft constraints of the segmentation determined by regional and boundary properties of A represented by cost function:

$$E(A) = \lambda \cdot R(A) + B(A), \qquad (1)$$

where $R(A)$ represents regional properties [1], which determine individual penalties for assigning pixels to the "objects" and "background".

Regional properties can be represented by:

$$R(A) = \sum_{p \in P} R_p(A_p), \qquad (2)$$

where $R_p(A_p)$ represents the *t-link* value of the edge that connects pixels with terminals.

Boundary properties [1] are determined by $B(A)$ and can be interpreted as penalization for discontinuities between two neighboring pixels p and q. This boundary property can be represented by:

$$B(A) = \sum_{\{p,q\} \in N: A_p \neq A_q} B_{\{p,q\}} \qquad (3)$$

Coefficient λ demonstrates the importance of regional properties against border properties.

The graph created like this consists of vertices and edges. The vertices [4] consist of all the pixels in the image and the terminals defined by the user. The edges [4] consist of all the edges including *n-link* edges that connect all neighboring pixels and *t-link* edges that connect all the pixels with the terminals.

The cut in the graph is located on the border between the object and background in the location of the maximum flow in the graph. It results is separating the object and background. This flow is determined using an algorithm to determine the maximum flow in the graph [2].

III. EXPERIMENTS

The final segmentation can be affected by coefficients changing the cost function. The following experiments will show the impact on differently set coefficients of the parameters on final segmentations. This experimental segmentation is not focused on the best result but the impact of parameters on the segmentation.

Graph cut segmentations are initialized by determining several pixels identifying the object in the image Fig. 1 [5] and several pixels identifying the background in the image.

The parameter $R_p(A_p)$ is defined as the difference of pixel value and terminals identifying the "object" and "background". By changing the value of the coefficients λ the impact of this parameter on the resulting segmentation is experimentally verified.

The parameter $B_{\{p,q\}}$ is defined as:

$$B_{\{p,q\}} = \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \qquad (4)$$

where σ is a coefficient defined by the user. By changing the value of the coefficient σ the impact of this parameter on the resulting segmentation is experimentally verified.

Fig. 1 Input image

Initialization is based on labeling the object and background of segmentation in Fig. 2. This initialization determines the object and background terminals for Graph cut segmentation.

Fig. 2 Initialization of Graph cut

The resulting segmentations are achieved by adjusting the coefficients for calculating regional and boundary properties of Graph cut segmentation. By changing the coefficients σ at zero values of coefficients λ the following segmentations are achieved as Fig. 3, Fig. 4 and Fig. 5 show.
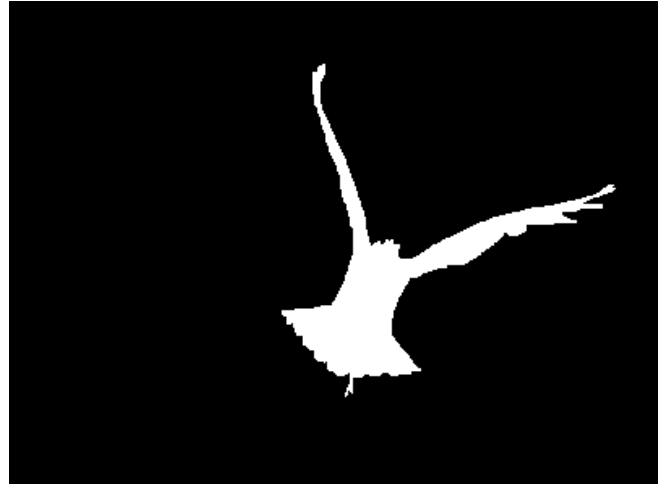
Fig. 3 Graph cut segmentation with $\lambda = 0$ and $\sigma = 0.1$
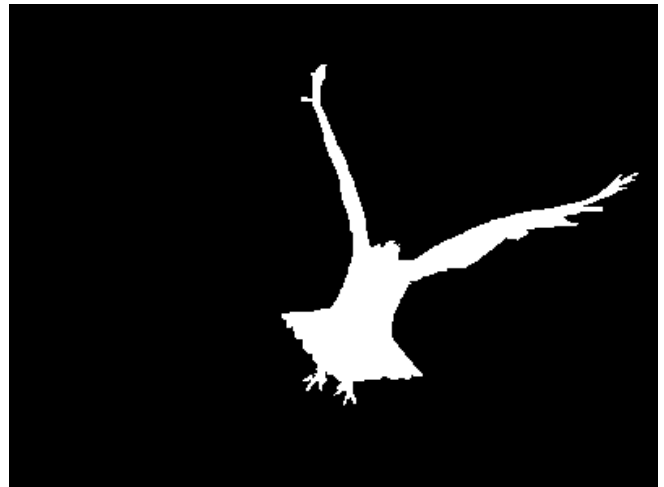
Fig. 4 Graph cut segmentation with $\lambda = 0$ and $\sigma = 0.5$
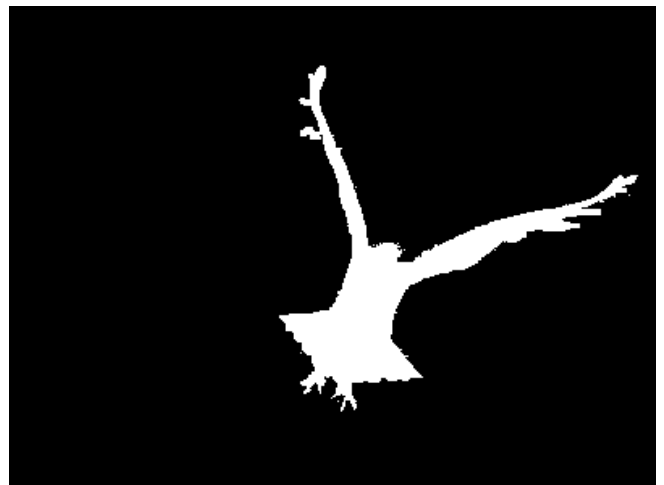
Fig. 5 Graph cut segmentation with $\lambda = 0$ and $\sigma = 1$

Changing only the σ coefficient results in more homogeneous segmented objects but near the borders they are not very sharp [7]. A too high value of σ will cause the objects on the border to intervene into the background of the segmentation Fig. 5. Unassigned pixels at the border with pixels assigned to the object or background are labeled as the object or the background although the value differences of the two adjacent pixels $(I_p - I_q)$ are bigger.

Fig. 6 and Fig. 7 show segmentations achieved by changing

the coefficients λ with previously used values of σ coefficients.

Fig. 6 Graph cut segmentation with $\lambda = 1e^{-5}$ and $\sigma = 1$

Fig. 7 Graph cut segmentation with $\lambda = 1e^{-5}$ and $\sigma = 0.1$

Fig. 8 Graph cut segmentation with $\lambda = 1e^{-4}$ and $\sigma = 0.1$

As a result of only a small change of the λ coefficients , the segmented objects near the border are sharper and do not intervene into the border to the background of the segmentation Fig. 7. If the coefficient λ is used, the coefficient σ must be changed to lower value. A too high value of σ will cause, that objects will be less homogenous

Fig. 6.

The difference between Fig. 7 and Fig. 8 clearly indicates that a little change of the coefficient λ has an impact on the resulting segmentation. Since initializing the object, the bird's beak is not labeled in the image; the correct result should be in Fig. 8.

IV. CONCLUSION

The aim of this article was to show the impact of the parameters after changing their coefficients on the resulting segmentations. The article provides visual evidence-based guidelines for determination of the initialization coefficients in order to achieve better segmentation results applicable for further processing [6]. Experiments show the importance of properly adjusted coefficients for parameters of Graph cut segmentation technique for the final segmentations.

REFERENCES

[1] Y. Boykov, G. Funka-Lea, "Graph Cut and Efficient N-D Image segmentation", *International Journal of Computer Vision*, vol.70, no.2, pp.109-131, 2006.
[2] Y. Boykov, V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization In Vision", *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no.9, pp 1124-1137, September 2004.
[3] Ch. Wang, L. Guan, "Graph cut video object segmentation using histogram of oriented gradients", *In ISCAS 2008: IEEE International Symposium on Circuits and Systems*, pp.2590-2593, May 2008.
[4] P. Karch, I. Zolotová, "Interactive Contour Deformation of an Object Using Graph Cut", INES 2010: 14th International Conference on Intelligent Engineering Systems, pp 227-230, May 2010.
[5] [Online]. Available: http://ivrgwww.epfl.ch/supplementary_material/ RK_CVPR09/HTML/SalientObjectSegmentation.html.
[6] S. Šimoňák, B. Sobota, "Image segmentation a possible input process for virtual reality objects creating", *In: Informatika a algoritmy '99*, pp.31-34, Košice, 1999, ISBN 8088941059.
[7] L. Muzikářová, M. Franeková, "Teória informácia a signálov". EDIS - ŽU, Žilina, 2009, ISBN 978-80-554-0075-4.

Improving the Accuracy of Classification and the Generalization Ability of an ARTMAP Neural Network

Peter SMOLÁR, Jaroslav TUHÁRSKY

Center for Intelligent Technologies, Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic, www.ai-cit.sk

{peter.smolar, jaroslav.tuharsky}@tuke.sk

Abstract—This paper proposes modification of a neural network (NN) of type Membership-function (MF) ARTMAP. Our approach generated from firstly, observing and analyzing the training phase of the network and secondly, from testing it during the object classification process. The first proposal of the modification is associated with the minimization of number of created fuzzy clusters. The second suggestion consists of the mathematical model of fuzzy relation, by which the membership of the input sample to the fuzzy cluster is calculated. As results we expect the accuracy improvement and generalization ability of the network and also acceleration of the testing phase.

Keywords— Adaptive Resonance Theory (ART), ARTMAP, Machine Learning, Probability Estimation.

I. INTRODUCTION

Hand in hand with the boom of the application potential of the information technology, raises demands for fast processing of large amounts of data. That is the reason why the methods used for obtaining relevant information asked by customers are under pressure.

Artificial intelligent (AI) methods are one of the possible ways how to handle with data. Neural networks, among others AI techniques, posses effectives features: the ability to adapt to dynamically changing problems, the ability to learn and also to handle uncertainty.

This paper introduces one of the type of NN, the network class ART (Adaptive Resonance Theory) and specifically ARTMAP. The approaches based on ART or ARTMAP technology are rather flexible and much more suitable for these situations when class consists of various clusters in the feature space.

This paper is organized as follows: Section II theoretically talks about the ART-like neural networks. The basic description of MF-ARTMAP is given. Section III contains our proposal of the improvement of this type of NN. Section IV presents our solutions for the minimization of fuzzy clusters.

II. NEURAL NETWORKS OF TYPE ARTMAP

Adaptive Resonance Theory (ART) was developed by Grossberg and Carpenter [6]. It was biologically inspired, trying to imitate how the brain processes information –analysis of human cognitive information and stable coding in a complex input environment [7]". Simply described, the pattern matching process compares the current input with a learnt category representation. Then this process leads to a resonant state (focusing attention and triggering category learning) or to a self-regulating parallel memory search leading to a resonant state. The important feature of ART NN is that if an established category is selected, the category´s representation may be refined to incorporate new information from the current input. In case that the search ends by selecting an antecedently untrained node, the ART NN establishes a new category.

ART NN has been used in a wide spectrum of application fields, including problems of learning and classification, recognition and prediction. Different types of ART NN have been developed: the first one by Grossberg was an unsupervised learning system used for categorize binary input patterns; the next ones were extended to categorize also the analog input patterns.

Neural networks of type ARTMAP, known also as predictive ART, have supervised learning structure, with ability to self-organize arbitrary mappings from input vectors representing features such as spectral values and terrain variables, to output vectors, representing predictions such as vegetation classes in a remote sensing application. If fuzzy ART units are used, we talk about Fuzzy ARTMAPS. It calculates the membership function of the point from the feature space to the fuzzy class. MF ARTMAP is based on the assumption that data in feature space are organized in fuzzy clusters. Grossberg in [6] it describes as a synthesis of fuzzy logic and adaptive resonance theory. Figure 1 illustrates the basic topology of the MF ARTMAP NN and in figure 2 is illustrated membership function of fuzzy relation for one cluster.

III. PROBLEM ANALYSIS OF MF ARTMAP NEURAL NETWORK

Analyzing MF ARTMAP behaviors during the phase of learning and making the analysis of results of test process we found some "weak points" of the network. Some of these deficiencies can be summarized as follows:

1. Used function of fuzzy relation have not sufficient possibilities for a description a real shape of cluster resulting from the training set for a more complex shapes of fuzzy clusters.
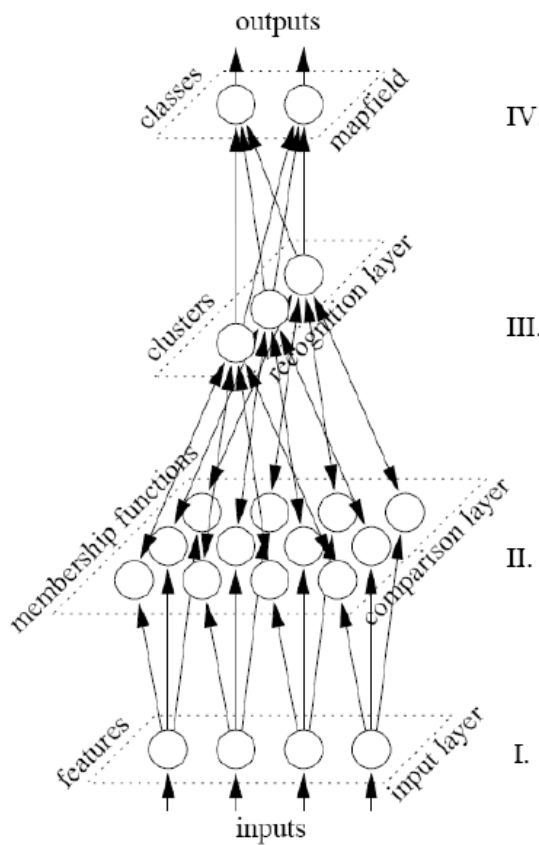
Fig. 1. Basic topology of the MF ARTMAP neural network [8].
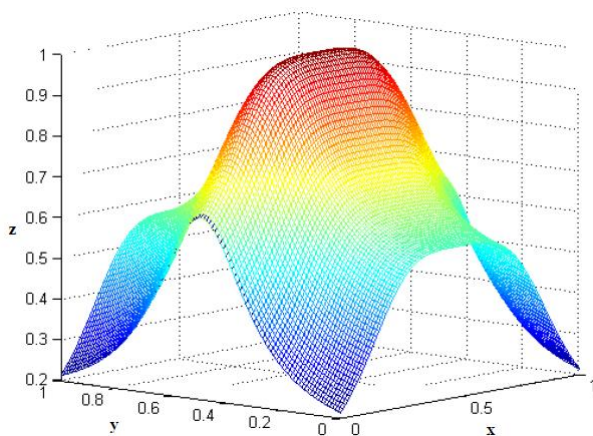
Fig. 2. Two-dimensional fuzzy relation [7]. Axes x and y represent dimensions of feature space. Axes z represents a membership of patterns to the fuzzy clusters.

2. During the classification process input samples have different values of membership to fuzzy classes, which do not correspond to actual values.

Based on these findings we propose three various modifications of MF ARTMAP neural network, which are expected to increase the accuracy of calculating membership functions of tested samples to fuzzy classes, increasing the accuracy of classification results, too.

IV. MINIMALISATION OF NUMBER OF FUZZY CLUSTERS

As is illustrated in figure "circle within a square" (Fig. 3) and „spiral" (Fig. 4), if the input samples are used in a random order to MF ARTMAP network, the network generates large amounts of new fuzzy clusters depending on the parameter threshold, although these fuzzy clusters are located side by side and belong to the same fuzzy class. Ideal creation of a fuzzy clusters for an image "in the square circle" in terms of number of clusters is illustrated by figure (Fig. 5), where the object of circle represents one fuzzy cluster.

A large number of fuzzy clusters cause that samples, which are on the edges of two clusters and belong to the same fuzzy class, have little value of membership function to the fuzzy class.

Proposed solutions to overcome this deficiency are as follows:

1. Some impact on learning has an order of training samples. Hence one of options is the selection of samples from training set to network according to the specified proximity (not exceeding the threshold parameter) to the fuzzy clusters that have already been created before. Then a new cluster is created only if no input sample can be classified to the already existing fuzzy clusters.

2. Other variety of modification is to start learning algorithms several times, but with a different sequence of input samples. Next step is unification of fuzzy relations from previous cycles of algorithm. This will increase the accuracy value of membership function of input sample to class.

3. Creation of next layer of learning algorithm in MF ARTMAP network, which will detect in the end of learning process, whether surrounding fuzzy cluster belong to the common fuzzy class or not. Then, based on fuzzy cluster environment, this cluster will "absorb" the surrounding clusters, if these fuzzy clusters belong to a same fuzzy class.

Lower number of fuzzy clusters cause faster running of algorithm and more accurate determination of membership of the sample to the class.
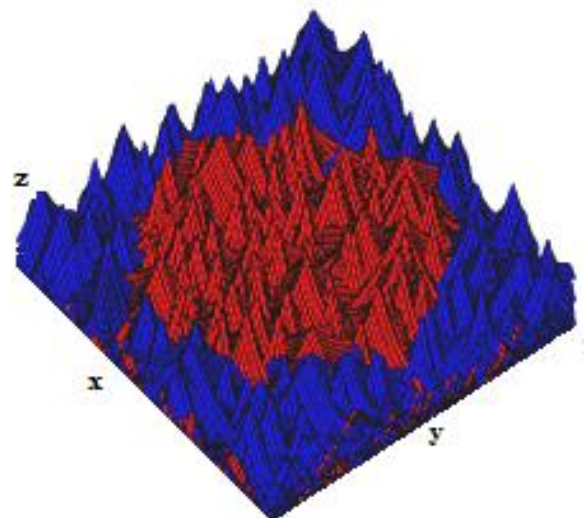
Fig. 3. Picture "Circle within square". There is shown a big amount of small clusters. It causes worse classifications results [7].
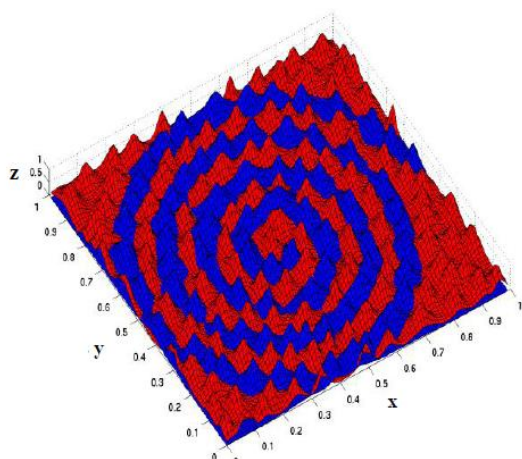
Fig. 4. Picture „spiral" with a big amount of small clusters [7].
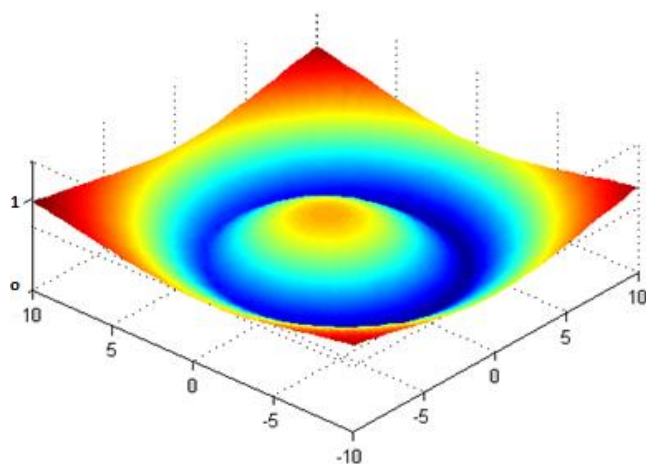
Fig. 5. Ideal adjustment of the fuzzy relation of the MF ARTMAP network from the point of view of number of fuzzy clusters created on the tested object which is Circle in Square. The class of Circle is

V. Conclusion

The main improvement expected from the proposed modifications is the increase of the accuracy of value of the membership function of input samples to fuzzy classes, by reducing the number of fuzzy clusters.

In our future work we would like to focus on further analysis and comparison of our proposed modification with the conventional MF ARTMAP network and with similar types of neural networks.

Acknowledgment

References

[1] A. Bodnárová. Metódy zhlukovania dát s využitím neurónových sietí typu ART. Diplomová práca, Technická univerzita v Košiciach, 2006.

[2] G. A. Carpenter, S. Grossberg, The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network, Computer, 21(3): 77-88, 1988

[3] G. A. Carpenter, S. Grossberg, and J. H. Reynolds. ARTMAP: Supervised Real-time Learning and Classification of Nonstationary Data by a Self-organizing Neural Network. Neural Networks, 4(5):565–588, 1991.

[4] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification, Second Edition. John Wiley & Sons, 2000.

[5] R.P.W. Duin and E. Pekalska. Open Issues in Pattern Recognition. In Computer Recognition Systems: Proceedings of 4th International Conference on Computer Recognition Systems CORES'05 (Advances in Soft Computing), pages 27–42, Secaucus, NJ, USA, 2005. Springer-Verlag New York, Inc.

[6] S. Grossberg. Adaptive Resonance Theory. Technical Report CAS/CNS-2000-024, Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems, September 2000.

[7] M. Hric. Inkremenálne učenie na báze ART neurónových sietí. Dizertačná práca, Technická univerzita v Košiciach, 2003

[8] A. K. Jain, M. N. Murty, and P. J. Flynn. Data Clustering: A Review. ACM Comput. Surv., 31(3):264–323, 1999.

[9] N. Kopčo, Biologicky motivované neurónové siete, Písomná časť dizertačnej práce, 1997.

[10] P. Maliňák, Inkrementálne učiace algoritmy pre rozpoznávacie systémy, Písomná práca k dizertačnej skúške, 2008.

Intrusion detection system architecture based on dataflow graphs

Eva DANKOVÁ, Peter JAKUBČO, Marek DOMITER

[1]Dept. of Computer and Informatics, FEI TU of Košice, Slovak Republic

eva.dankova@tuke.sk,peter.jakubco@tuke.sk, marek.domiter@tuke.sk

Abstract—an intrusion detection system is today very important area in network security. IDS which are today used have many lack of, so it is needed to design IDS, which would be more flexible and complex. This paper handles about design of IDS system which is based on dataflow graphs. In first part are analyzed IDS systems to show their advantages and disadvantages. Second part handles about classification scheme for the intrusion marks, which is necessary for the design.

Keywords—dataflow graph, attack, intrusion, anomaly, sign

I. INTRODUCTION

Intrusion detection system is relative a young technology, which widespread intrusion detection study started in 1980 Conclusion of this study was large sphere of strategies and solutions for intrusion detection goal realization.

Beginning of intrusion detection brought also complication, because between theoretical and practical ambit still exists an abyss. This situation built all arts of proofs for examined and developing products in this area. There are many tryouts to define terms on the fly and to develop adequate solutions, which cooperates with other part of the security system or with control infrastructure. Next important tryout is the requirement that preferred solution should solve all problems aside from argument availability.

This disadvantage will be solve depending up to user requirements and up to growing financial facilities reserved for study in this area. Importance of intrusion detection in defensive information war is clear and therefore was reserved additional financial facilities for this area.

II. COMPUTER SYSTEM SECURITY

From practical view is most popular and simplest definition of secure system formulation of Simson Garfinkel:

"Secure computer system is dependent on behavior, which is assumed. " [1]

In principle computer system is secure, if it is possible to trust it. Level of this trust in computer system indicates trust level of assumed computer behavior, which can be in a form of security policy. Analysis of security functions can be valuable process part in security planning. Powerful security policy has to cover risk avoiding, discouraging, prevention, detection and recovery. Computer security definition [2] is derived from level of realization possibilities, confidence, integrity and availability in computer system. Confidence requires information availability only for authorized user. Data integrity can be understood as condition, which ensures that data stay unchanged.

A. Security threats

With the growing requirement on information accessibility, increase also quantity of data saved in computer system. Growing connectivity brought the possibility to access the quantum of data and information, but also this access allowed from different places in the network.

Threat is possible to define as different situation or events, which have potential to damage system. Damage can be in a form of data detection, destruction or modification, or it can disable access to data or system sources. Threats can be divided in categories:
- External intrusion – not authorized user
- Internal intrusion – authorized user, who overstep his legal rights.

This art of intrusion can be also classified as followed:
- Masked user – user, who use by intrusion authentication data of other users.
- Illegal user – user, who gets around audit and control proceeding.
- Violation – authorized user, who overstep his privileges.

Intrusion detection is base upon models, in which is used "intrusion" to describe all international corruptions of system security policy. Today, if the only prevention is authentication, it is not a problem for attacker to gain authentication data from other user, or to avoid the whole authentication process. With technology development accrue also knowledge about computer system, based on witch attacker could inquire system inadequacies. These inadequacies can be used by intrusion. This attacks are often well masked, so it is not easy to detect them in often used computer systems.

B. Intrusion detection

Regarding to today's conditions a scale is necessary to determine security level, because of shorter executing period, and also because of bigger software faultiness.

One possible solution is intrusion detection, which simultaneously represent the last level in common model for system security. Intrusion detection system (Fig 1.) is not

used only for successful intrusion detection, but also for attempt intrusion monitoring.
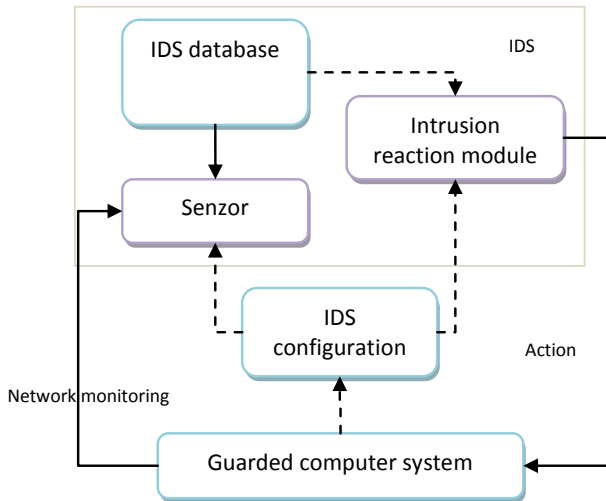
Fig. 1. Intrusion detection system [3]

Sensor task is to filter information and to throw off all irrelevant data acquired from event folder, which are connected to guarded system. Based on this filtration can be detected abusive activities. For this purpose is used detection policy database, which consists of: sign-based thrust, normal behavior, profiles, needed parameters. In addition this database contains IDS configuration parameters, communication approach with reaction module. Sensor has also its own database, which consists of complex dynamic history of potential intrusions.

C. Principle of intrusion detection

In intrusion detection process after defining intrusion types is also important detection method. By monitoring intrusion with use of detection method is system behavior inquired, if there are some errors flags or some other security breaks. In intrusion detection most methods include also defect detection [3], anomaly detection [3], or combination of both.

- Misuse detection – Detection is oriented in monitoring behavior, which can be defined as faulty. During detection are filtered data flows, where by is looking for activities signs, which correspond to known attacks or to other security breaks. Failure detection use technique based on signature comparison. Signatures used by compare are known and reference to problems. The most actual commercial system for intrusion detection use also technique based on failure detection.
- Anomaly detection – Detection is oriented in monitoring behavior, which can be qualified as unusual or rare. It analyzed system data flows with use of statistic method to find signs or activities, which can be classified as abnormal. This approach reflex view of some scientist, that intrusions are subsets of anomaly activities.

With conjunction both methods, it acquire important advantages. Anomaly detection allows system to find new or unknown attacks. Failure detection protects anomaly detection integrity, so that attacker can not modify detector so that it would sense attack behavior as normal.

III. INTRUSION DETECTION MODEL

Intrusion detection model, which is independent from system, was designed by Dorothy Denning [4] in 1987. this model is type of monitored intrusion entries and specifications. Short description of global model is useful in context to specification system models for intrusion detection. This global model is still exact in describing architectures of many actual systems. On following figure (Fig 2.) is image global model for intrusion detection, where event generator is global.
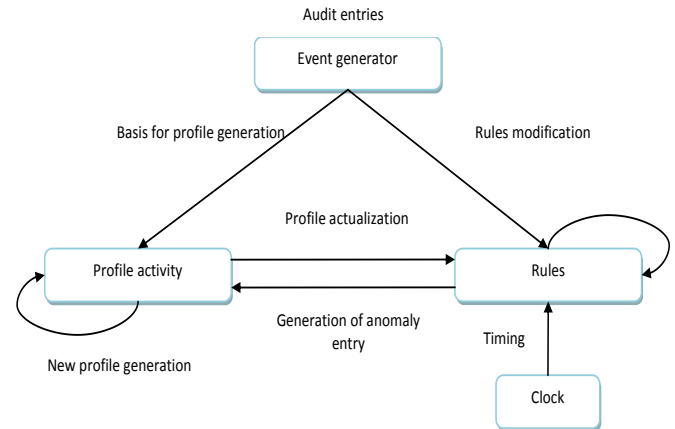
Fig. 2. Intrusion detekcion model [3]

Activity profile is global intrusion detection state, which include variables for system behavior computing with use of predefines statistic measures. These indicators are intelligent variables, what mean, that every variable is connected to exemplary specification used to filter event records.

Suitable records provide data for their value actualization. Every variable is connected to one statistic measure built in system, and is responsible for actualization of their state based on information contained in event records.

Activity profile can dynamically create new profiles for new created subjects based on pattern. If there are added new users or folders in the system, these patterns build new profiles for them. Activity profile can also generate anomaly records in case, that some statistic variable acquires anomaly values.

Rules introduce global derivation mechanism and use event records, anomaly records and expiration time to control activities of other components and actualization their state.

IV. DESIGNED SOLUTION METHOD

Methods applied to solve the problems in intrusion detection are:

- Analysis and classification of intrusion signs
- Planning method of partly ordered events.
- Potting method of input data flow with use of dataflow graphs.

System intrusion is represented by event sequences, which are characteristic from view of intrusion signs and context facilities. If there is equally fine grained logging system, which is noting individual signs and their context, it is possible to detect these signs. Mutual relations between low and high leveled events are described with use of global abstract structure. This structure integrates various input data from different input logs.

Planning method of partially ordered events is working over an input data set about intrusion. Individual elements from the set represent intrusion signs and their facilities are given through mutual disposition of some elements from input set. Disposition operation represents binary operation denotative presence one event in relation to other in sequential order. One of the main intrusion detection problems is limitation of variability detection ability from the same attack. The same attack is possible to represent with other event sequence.

One of the planning facilities partly ordered events is possibility to create an event sequence plan. This plan describes when some events are executed independent from each other. Conclusion from the detection view is plan in a form of dataflow graphs in a form of acyclic oriented graph. Single places in a graph represents intrusion sign context and edges represents context conditions of overpass between single places.

Signs description represents parallel system – events can be executed independent from each other. The core of designed detection is built from several intrusion descriptions in form of dataflow graphs. Dataflow graphs are adequate language suitable for system modeling with inside parallelism. This parallelism appears asynchronous and parallel behavior.

Conclusion of planning dataflow graphs is resultant plan transformation in to formal specification by dataflow graphs. Final IDS system architecture represents dataflow graphs set, which represent intrusions.

V. Intrusion signs classification

Intrusion signs hierarchy is possible to derivate based on relations between particular high leveled events. An advantage of this hierarchy is that it allows signs classification based on common characteristic. These characteristics are abstracted from any kind of specification and this signs detection, what serves conceptual advantages for understanding intrusion signs.

Classification of concrete signs in to given class express structural relations between high leveled events. Designed hierarchy classified intrusion signs as autonomous threat. Dividing software effects is in software engineering still alive, and therefore is necessary to create categories of these defects. If a defect occurs often or is dominant, than it is possible based on education, training, documentation, inspection and other methods to reduce occurrence of this defect to minimum. In a fall that there is no occurrence statistic, than it is not possible to think about timely detection and elimination of this defects. The main goal is prevention. Through defect problematic study in software systems is possible to design techniques for defects reeducation.

Designed scheme includes 4 categories:
- Existence signs ,
- Sequence signs,
- Partially ordered signs,
- Other signs.

In this scheme a category of higher level roof lower categories in terms of signs. This signs can be represented through this category. Exact limitation for plotting is possible to create by creating instance of this category. Object, which should be created, require exact defined structure of high leveled events in meaning of low layered events.

By defining high leveled event structure in a more complex form has as result signs sift from higher to lower level. High leveled events serve for smoothing differences between log records. Through this smoothing partially intrusion signs stay untouched, if they are described by high leveled events. Hierarchy was designed based on standard logging records, where was deliberate, that system operation calls implementation of reading and writing represent in logging records independent operations for each attribute specified for this callings.

VI. Conclusion

Goal of designed classification is intrusion signs categorization, which are independent from plotting computing model of the signs. Therefore classification serves as basis for computing model derivation to find intrusions in a system. Using this classification can be developed in two planes. On individual categories it is possible to look as on disjunctive sets, where each includes own methods for intrusion sign detection. Conclusion of this paper is creating a consolidated method for intrusion detection. In case that it is possible to consider operating systems, which have the same defects, than existence of this errors in logging records is similar. An example can be side-run attack, which is founded in most operating systems, whereby it has the same global character with possibility of representation based on sign sequence.

References

[1] D. Gollmann," Computer security." John Wiley & Sons, Inc., New York, NY, USA, 1999.
[2] D. Russell and Sr. G. T. Gangemi. "Computer security basics.", O'Reilly & Associates, Inc., Sebastopol, CA, USA, 1991.
[3] R. Gurley Bace." Intrusion detection". Macmillan Publishing Co., Inc., Indianapolis, IN, USA, 2000.
[4] D. E. Denning. "An intrusion-detection model.", IEEE Trans. Softw. Eng., 13(2):222{232, 1987.
[5] L. Vokorokos, "Data Flow computing model: application for parallel computer system diagnosis." Computers and artificial inteligence, Vol. 20, No. 4/2001, SAP, ISSN 0232-0274.
[6] L. Vokorokos, A. Baláž, and M. Chovanec, "Intrusion detection system using organizing map.", Acta Electrotechnica et Informatica, pages 81-86, 2006 Februar

Language Patterns:
Reusable Elements of Computer Language Design

Miroslav SABO

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

miroslav.sabo@tuke.sk

Abstract—**The current adoption of domain-specific approach to software development naturally leads to increased demand for development of domain-specific languages (DSLs). Even though the notations of DSLs share some common features, they are still defined manually and from scratch for each DSL. This paper presents a novel notion of language pattern, as a formal way of capturing the recurring specifications in computer language design. The main goal is to facilitate the process of a notation specification by automating the specification of common features often used in various programming languages. The application of language patterns is presented in the context of method for example-driven DSL notation specification. By using the method, domain expert becomes an active member of the language development process, which is crucial for the creation of the successful DSL.**

Keywords—**concrete syntax, domain-specific language, language design, language pattern, notation**

I. INTRODUCTION

When planning a development of software solution in the specific narrow application domain, domain-specific language (DSL) is often considered as one of the options for its implementation [1], [2]. Despite its obvious benefits, such as expressivity in the concepts inherent to the domain, self documenting, domain-specific validation and others, DSL is usually not chosen due to high costs of its implementation.

Various solutions exist to enhance the process of DSL development, MetaEdit+ [3] and Visual Studio DSL Tools [4] for graphical languages, and EMFText [5], MPS [6], TCS [7] and Xtext [8] for textual languages. Each of them uses its own approach to define the notation of a language, but in all of them, the process is carried out explicitly by language engineer, sometimes in cooperation with domain expert.

It often happens that the particular notation, commonly known either to be required because of requirements on the ambiguity in a formal language, or as a natural representation for some language construct, has to be specified manually and repeatedly for each of its usage. The prominent examples are *keywords* before otherwise indistinguishable language constructs and *delimiter marks* as a way to separate the individual items in a list.

We aim for recognizing and automated handling of these cases. To achieve this, we propose a notion of *language pattern* as a reusable element of computer language design. These patterns would capture the knowledge of language expert and enable the assisted specification of formal language, even by a person less versed in this area. Another application would be the example-driven inference of a language notation from provided program samples.

The rest of the paper is organized as follows. Since the proposed notion of language pattern is defined in the context of model-driven language development, necessary rationale is described in detail in Section II. Section III introduces the notion of language pattern and its formal model. The classification of up to now identified and formalized language patterns is presented in this section as well. Section IV discusses the concrete example of application of Separator language pattern in example-driven method for DSL notation specification. The comparison with similar approaches to DSL development is briefly discussed in Section V. Finally, Section VI gives the conclusions of the paper.

II. MODEL-DRIVEN LANGUAGE DEVELOPMENT

Contrary to traditional approach based on using Extended Backus-Naur Form (EBNF) as a formal way to describe computer programming language, we are focusing on the description of a language using a metamodel [9], [10], [11], particularly Ecore metamodel [12]. According to Meta Object Facility (MOF) four-layer metadata architecture [13], Ecore and EBNF reside at the topmost M3 layer, and serve as meta-metamodels for defining metamodels at M2 layer. Examples of such metamodels would be UML or OWL metamodels for Ecore, and Java or C grammars for EBNF. These metamodels at M2 layer provide concepts for defining models of the real world at M1 layer. Finally, at M0 layer reside things from the real world.
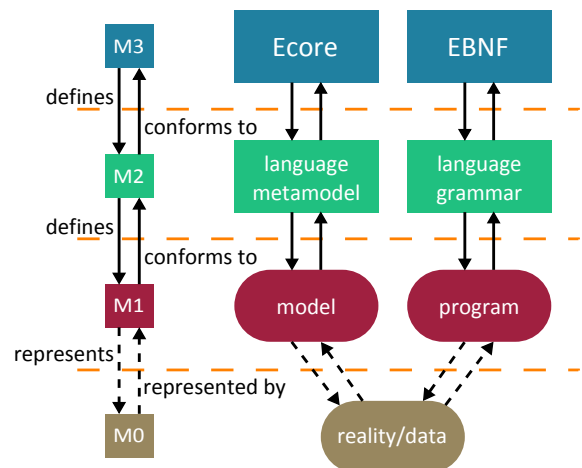
Fig. 1. Comparison of metamodel-based and grammar-based modeling spaces in a four-layer metadata architecture.

In the context of model-driven language development and method for example-driven DSL notation specification (which will be in more detail described in Section IV), the architectural overview of artifacts used throughout this paper is depicted on Fig. 2. The left branch illustrates *ProgSample* modeling language, the purpose of which is to represent the provided DSL program samples in a tree form composed of specific nodes (such as word, whitespace, delimiter, comment, block, etc.). The right branch illustrates the metamodel of an actual DSL which is being specified in a method. From the perspective of syntaxes, metamodel on the right side represents the abstract syntax of the DSL, while program sample (or its ProgSample model) on the left side represents its concrete syntax.
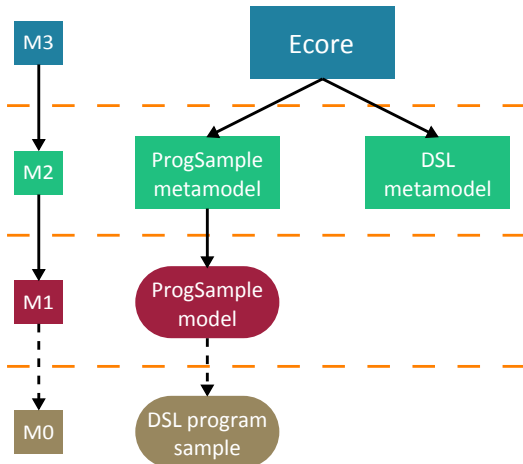
Fig. 2. Architectural overview of the artifacts used in example-driven method for DSL notation specification.

A. ProgSample Language

ProgSample language is a DSL for tree representation of textual files. In the context of method for example-driven DSL notation specification, the textual files are samples of DSL programs as they would have been written in a DSL being specified.
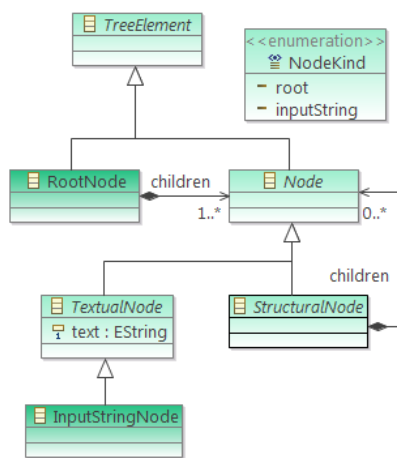
Fig. 3. Core metamodel of *ProgSample* language for defining tree models of program samples.

The special feature about this language is that its metamodel is not defined as a monolithic structure, but rather as a

composition of nodes attached to the metamodel core. In its simplest form, with no extending nodes, the core metamodel is depicted in Fig. 3. According to this metamodel, textual file can be modeled as a tree consisting of the *RootNode* root with single *InputStringNode* node holding the whole text of a file. This is just the most trivial case and indeed, such model does not add any structural information, comparing to the flat textual representation of a file. However, this initial tree model serves as an input for the method for notation specification, and by application of *formatting language patterns* (Section III-A), it can be transformed into more complex form.

III. LANGUAGE PATTERNS

The notion of *language patterns* in the area of computer language development has been inspired by the notion of design patterns in the area of object-oriented software [14]. In analogy with design patterns, language patterns capture the language design experience in a form that people can use effectively.

From the point of view of which syntaxes are addressed by the language pattern, we distinguish two kinds – *formatting* and *mapping* language patterns (Tab. I).

A. Formatting Language Patterns

Formatting language patterns address only that part of *concrete syntax* design which does not directly concern the representation of concepts from abstract syntax. The examples would be definition of whitespace characters, starting/ending marks for one line/multiple lines comments, starting/ending marks for blocks or structuring to blocks by indentation. Usually, it is possible to identify all of these features of the notation simply by giving a short look at a piece of code. It is so because these features follow the patterns commonly known and observed in various programming languages.
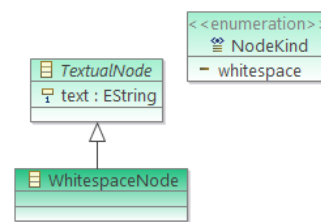
Fig. 4. Extending *ProgSample* metamodel elements, introduced by *Whitespace* language pattern.

These patterns are formally defined in terms of ProgSample metamodel elements as transformations of ProgSample models. It often happens, that the transformation introduces new kind of nodes to the model. In such cases, besides the definition of a transformation, formal model of the pattern also contains definition of the extending metamodel elements and their connection to the ProgSample core metamodel (Fig. 4). There are only two possible extension points – *TextualNode* and *StructuralNode* nodes, one for the new kind of nodes for holding textual information and the other for holding structural information. The new kind of extending node must be recorded in the *NodeKind* enumeration as well.

TABLE I
COMPARISON OF FORMATTING AND MAPPING LANGUAGE PATTERNS.

	Parsing language patterns	Mapping language patterns
Purpose	construction of *ProgSample* models representing DSL program samples	definition of abstract to concrete syntax mapping by annotating the DSL metamodel
Artifacts	program samples program sample models	program sample models DSL metamodel
Extending *ProgSample* metamodel	Yes	No
DSL metamodel template	No	Yes
Result of application	modification of program sample models	modification of program sample models, annotation of DSL metamodel
Examples	Whitespace Tabulator Word Comment Block	Keyword Concept Separator Range Optional Before After

TABLE II
LIST OF UP TO NOW IDENTIFIED AND FORMALIZED LANGUAGE PATTERNS.

Language pattern	Intent	Type
Whitespace	definition of whitespace characters	F
Tabulator	definition of tabulator characters	F
Word	tokenizing text string words delimited by whitespaces	F
Comment	definition of comments	F
Block	dividing the sequence of elements into blocks	F
Keyword	keywords in DSLs are often defined as names of concepts used in the abstract syntax (domain model) of a language	M
Concept	notation of concept usually starts with a keyword defining the name of a concept	M
Separator	in the notation of list, items are usually separated by some delimiter mark	M
Range	multiplicity of the concept	M
Optional	optional occurrence of a concept	M
Before	prefix notation	M
After	postfix notation	M

B. Mapping Language Patterns

Mapping language patterns address both, *abstract* and *concrete syntaxes*, as they concern how the concepts from abstract syntax are realized in a concrete human-usable notation. From the technical point of view, they express the mapping of abstract syntax to concrete syntax. Their purpose is twofold:

1) control the specified notation, that it satisfies requirements on being computer processable (problem of ambiguity in a language)
2) capture the widely known and accepted notations for particular language constructs, often used throughout various programming languages

The example of a mapping language pattern would be the *Separator*. This pattern captures widely accepted notation for listing of multiple items of the same kind, where individual items are separated by some particular delimiter mark like comma or semicolon. The formal model of mapping language patterns is defined on M2 layer of the metadata architecture, in terms of Ecore meta-metamodel. This model prescribes the template of a structure of elements which constitute the language construct being addressed.

Fig. 5. DSL metamodel template defined in Separator language pattern.

The template of *Separator* language pattern, as depicted on Fig. 5, defines the structure of listing language construct. It is composed of two concepts, one for defining container and other for defining items, and a containment relationship among them. If such structure is recognized in the metamodel of a DSL being specified, then accordingly to the common experience, its notation should follow some predefined rules, such as using of delimiter marks in this particular case. This is specified by assigning of annotation to this structure in a DSL metamodel.

IV. EXAMPLE-DRIVEN METHOD FOR DSL NOTATION SPECIFICATION

Language patterns play the key role in a method for example-driven DSL notation specification [11], [15], [16]. In contrast to traditional approach to computer language development where specifications of abstract and concrete syntaxes are mixed together in a grammar, we advocate the approach of separate specifications. Another advantage of the method is that domain expert, as a person most versed in the domain of a DSL, can be incorporated more into the process of its specification. Overview of the method is illustrated in the Fig. 6.

The method starts with domain analysis, conducted together by domain and language experts. The output of this process is a domain model which represents the abstract syntax of a DSL (metamodel) being specified. This model is constructed by using Ecore (Fig. 7).

Concrete syntax is specified solely by domain expert, as she is the most competent person for the task. Since domain experts usually do not possess sufficient knowledge to express the notation formally, the proposed method takes this fact into consideration and is designed to allow domain expert define the notation by writing the examples of programs as they
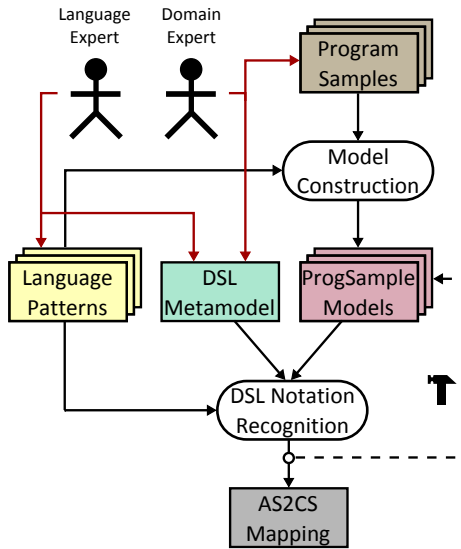
Fig. 6. Overview of the method for DSL notation recognition from provided program samples.
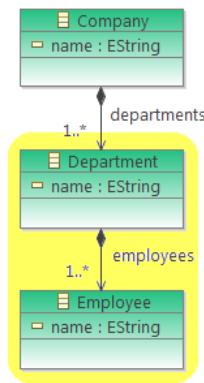
Fig. 7. Example of metamodel of a simple DSL for defining personal structure of a company.

would have been written in a desired DSL (Fig. 8). Third input to the method is a set of language patterns, formalized in models, as discussed in Section Section III.

Fig. 8. Example of the DSL program sample.

The method consists of two consecutive processes. Firstly, the provided DSL program samples are transformed into *ProgSample* models by using *formatting language patterns*. After models have been created, they serve, together with DSL metamodel, as inputs into the process of DSL notation recognition driven by *mapping language patterns*. The method is in detail described in [16]. Recognition of the *Separator* pattern (Fig. 5) is illustrated in the Fig. 7 and Fig. 8.

V. RELATED WORKS

Considering the model-driven DSL development, three prominent technologies are closely related to our approach of using the language patterns.

TCS [7] is a DSL for the specification of Textual Concrete Syntaxes in model engineering. Similar to the method proposed in this paper, TCS allows the specification of a notation separately from the abstract syntax (domain model) without impacting each other's structure. However, the notation is defined rather through templates extending the concepts from domain model, than by inferring from examples of the programs.

The other tools, EMFText [12] and Xtext [8], are similar to our method in a way that domain model in both tools is defined using the EMF platform. The difference lies again in the way of the definition of a notation for DSL. EMFText, as well as Xtext, use their own templating mechanisms and user has to define them explicitly.

VI. CONCLUSION

Capturing of common notation style of particular language constructs and its formalization in a form of language patterns is an unexplored topic in the area of computer language development. This paper elaborates on this novel idea and briefly discusses its application in model-driven language development, particularly example-driven DSL notation recognition. Paper presents detail description of the formal model of the notion of language pattern and provides listing of patterns identified and recognized up to not as well.

ACKNOWLEDGMENT

REFERENCES

[1] S. Kelly and J. P. Tolvanen, *Domain-Specific Modeling*. Wiley-Blackwell, 2008.

[2] C. Simonyi, M. Christerson, and S. Clifford, "Intentional software," in *Proceedings of the 21st annual ACM SIGPLAN conference on Object-oriented programming systems, languages, and applications*, ser. OOPSLA '06. New York, NY, USA: ACM, 2006, pp. 451–464.

[3] MetaEdit+, http://www.metacase.com.

[4] Visual Studio DSL Tools, http://www.domainspecificdevelopment.com.

[5] EMFText, http://www.emftext.org.

[6] Meta Programming System, http://www.jetbrains.com/mps.

[7] F. Jouault, J. Bézivin, and I. Kurtev, "Tcs: a dsl for the specification of textual concrete syntaxes in model engineering," in *Proceedings of the 5th international conference on Generative programming and component engineering*, ser. GPCE '06. New York, NY, USA: ACM, 2006, pp. 249–254.

[8] Xtext, http://www.eclipse.org/Xtext.

[9] D. Djuric, D. Gasevic, and V. Devedzic, "The tao of modeling spaces," *Journal of Object Technology*, vol. 5, no. 8, pp. 125–147, 2006.

[10] A. Kleppe, *Software Language Engineering: Creating Domain-Specific Languages Using Metamodels*, 1st ed. Addison-Wesley Professional, 2008.

[11] J. Porubän, M. Sabo, J. Kollár, and M. Mernik, "Abstract syntax driven language development: defining language semantics through aspects," in *Proceedings of the International Workshop on Formalization of Modeling Languages*, ser. FML '10. New York, NY, USA: ACM, 2010, pp. 2:1–2:5.

[12] Eclipse Modeling Framework, http://www.eclipse.org/modeling/emf.

[13] Meta Object Facility (MOF) 2.0 Core Specification, http://www.omg.org/spec/MOF/2.0.

[14] E. Gamma, R. Helm, R. Johnson, and J. M. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional, 1994.

[15] M. Sabo, "Abstract syntax driven concrete syntax recognition," *Journal of Information, Control and Management Systems*, vol. 8, no. 4, pp. 393–402, 2010.

[16] M. Sabo and P. Jaroslav, "Concrete syntax recognition using language patterns," in *Proceedings of CSE 2010 International Scientific Conference on Computer Science and Engineering*, 2010, pp. 101–108.

Lisp Syntax and its Alternatives

Sergej CHODAREV

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

sergej.chodarev@tuke.sk

Abstract—**Lisp is a family of programming languages, that have a lot of distinct features. One of them is a very minimalist syntax based on lists, where program is represented using basic data structures of the language. This provides flexibility to define new language constructs. This paper describes this syntax and presents a number of alternative syntaxes for the languages of Lisp family that was developed during the history of the language.**

Keywords—**Homoiconicity, Lisp, syntax, S-expression.**

I. INTRODUCTION

Lisp is one of the oldest programming languages still in use. Lisp introduced a number of new concepts into programming and influenced further development of programming languages. One of its distinct features is its fully parenthesised syntax with infix operators.

```
(defun factorial (n)
  (if (<= n 1)
      1
      (* n (factorial (- n 1)))))
```

Figure 1.   Definition of factorial function in Common Lisp.

This syntax, known as S-expressions, is extremely simple, it is basically just a serialization of abstract syntax tree (AST). For this reason it can be said that Lisp has no syntax at all. All code is expressed using the same notation as used for expressing data structures. This property, where program code is represented using basic language data structures, is called *homoiconicity* [1].

This allows to manipulate code in the same way as data. One of the consequences of this is powerful macro system available in most of Lisp dialects, which is considered one of the most important Lisp features. Another advantage of S-expressions is their uniformity — user-defined constructs have the same form as build-in constructs. These properties allow language to grow with the help of users [2]. This also allows to use Lisp as a host language for custom domain-specific languages [3].

But at the same time, this syntax is very unusual. The usage of parentheses as the only way to group expressions and language constructs can make programs hard to read[1]. Prefix notation for operators is also inconvenient. For these reasons there was a number of efforts to create alternative syntax for Lisp. While being more convenient they tried to keep (at least partially) flexibility of S-expressions.

Syntax of Lisp and its dialects is great example of extensible syntax for programming language. For this reason it can be used as a basis for design of new extensible languages.

---

[1]It is also a reason for jokes, even the name of the language is sometimes explained as "Lots of Irritating Superfluous Parentheses" [4].

II. HISTORY OF LISP

Lisp was created in 1958 by John McCarthy. McCarthy explained early history of the language in [5]. Lisp was designed for artificial intelligence research, so it needed to make computations with symbolic expressions. It was decided to use nested lists for representing symbolic expressions. These lists were written in parentheses. This part of the syntax was called S-expressions (for "symbolic expressions") and was intended to represent data structures. Programs itself would be written using M-expressions (for "meta-expressions") with syntax similar to Fortran.

Using nested lists of symbols and other data values it was possible to express every symbolic expression and even Lisp programs itself. McCarthy introduced *eval* function, which would evaluate Lisp code represented as Lisp data structure. S. R. Russell implemented the *eval* function and so first Lisp interpreter was created. This caused that Lisp programmers started writing programs using S-expressions and M-expressions were not implemented and not even fully specified.

III. S-EXPRESSIONS AND M-EXPRESSIONS

S-expression is either an atom or a pair of other S-expressions enclosed in parentheses and separated by dot [6]. Pair fields are called *car* and *cdr*. Examples of S-expressions:

- `FOO` (single atom)
- `(X . Y)` (a pair)
- `((1 . 2) . (3 . 4))` (nested pairs)

Pairs can be used to represent lists by placing element of a list into first field and rest of a list into second field. For example `(1 . (2 . (3 . NIL)))`, where `NIL` represents empty list, is a list of three numbers. There is also an abbreviated syntax for lists, where list items are just enclosed in parentheses and separated by spaces. For example, previous list can be written as `(1 2 3)`.

Atoms can be symbols, numbers, strings, or other objects. Lisp symbols are unique strings and they have similar role as identifiers in other languages — values or procedures can be bound to them. But symbols can also be manipulated as data values.

M-expressions were supposed to be used for expressing programs manipulating with S-expressions. Definition of example function `ff` using M-expressions is on Fig. 2 (a). This function will find first atom in S-expression [7].

As can be seen from example, function arguments are enclosed in square brackets and separated by semicolons in M-expressions. There is also special syntax for conditional expressions.

M-expressions can be translated to S-expressions using simple rules. On Fig. 2 (b) is the same function represented using

```
a)   label[ff; λ[[x]; [atom[x] → x;
                       T → ff[car[x]]]]]

b)   (LABEL FF (LAMBDA (X) (COND
         ((ATOM X) X)
         ((QUOTE T) (FF (CAR X))))))

c)   (defun ff (x) (cond
         ((atom x) x)
         (t (ff (car x)))))
```

Figure 2.   Example function using M-expressions (a) and S-expressions (b, c)

S-expressions in Lisp 1.5 [7]. For comparison on Fig. 2 (c) is presented definition of similar function in modern Lisp dialect — Common Lisp.

When S-expressions are evaluated, atoms are evaluated to itself, except of symbols. Symbol is evaluated to the value bound to it. List is interpreted as function call – first item of a list is considered a function name and rest of a list is passed to function as arguments. If list contains nested lists, they are evaluated first.

IV. S-EXPRESSIONS IN MAIN LISP DIALECTS

There exist a lot of Lisp dialects and some of them introduce some modifications of S-expressions syntax. Most popular dialects today are Common Lisp [8] and Scheme [9]. Their S-expressions syntax is very similar.

Symbols are constructed from alphanumeric characters and some special characters like +*-, as long as they do not match syntax for numbers. This allows to use common operators as symbols and to use dash to separate words in long symbol names, e.g. `very-long-name`. Symbols can also contain escape sequences.

Numbers can contain sign, decimal point and exponent. Lisp also supports ratios, like `10/3`. *Strings* are enclosed in quotation marks and can contain escape sequences beginning with backslash. *Character* constants begin with sharp sing and backslash, e.g. `#\a`.

To avoid usual evaluation of lists and symbols, they can be *quoted*. A single quote character (') is used to quote S-expression[2]. For example `'(+ 1 2 3)` will not be interpreted as a sum of numbers, but as a list of symbol + and numbers 1, 2 and 3.

Similar feature is *quasiquotation*. It allows expression to be used as a template, where some parts should be replaced by the result of expression evaluation. Expression is quasiquoted if it is preceded by backquote character (`). Inside a quasiquoted expression, subexpressions preceded by comma will be evaluated and replaced with the result of evaluation. For example expression `` `(list ,(+ 1 2) 4) `` will be transformed into `(list 3 4)`.

Comments in most Lisps are started by semicolon and continue to the end of line. Both Common Lisp and Scheme also provide block comments in form `#| ... |#`.

Sharp sign (#) is traditionally used in Lisp to start some less common constructs. Its meaning depends on character that follows it. Besides chars it is also used to express vectors (lists with random access indexed by integers), like `#(1, 2, 3)`, or numbers with different radix, like `#b00011` for binary number.

[2]Single quote `'x` is actually just a shortcut for `(quote x)`.

Interesting feature of Common Lisp is that its syntax is defined using reader table. This table defines type of every character and can be modified by programmer. This provides additional possibility to extend the language. Character can be defined as a macro character. If such character would be read from input, associated function would be called, that would process following characters. Lists, strings, comments are actually implemented as reading macros in Common Lisp.

Scheme, compared to other Lisp dialects, allows to use square brackets instead of parentheses to mark lists. It also provides special notation for boolean values, like `#t`, and `#f`, and notation for complex numbers, like `3+4i`. Another interesting feature is possibility to comment out whole S-expression by prefixing it with `#;`.

A. Clojure

Clojure is a new Lisp dialect created in 2007. It runs on Java Virtual Machine and has good support for concurrent programming [10]. Its syntax is based on S-expressions but introduces several new elements:

- *Keywords* — separate data type similar to symbols, but evaluating to themselves: `:foo`
- *Vectors* enclosed in square brackets: `[1 2 3]`
- *Maps* — key-value pairs in braces `{:a 1 :b 2}`
- *Sets*: `#{:a :b :c}`
- *Regular expressions patterns*: `#"ab|bc"`
- *Metadata* — a map that can be associated with any object if written before it: `^{:a 1 :b 2} foo`

Other interesting feature is, that Clojure ignores commas and threats them as whitespace. This allows to use them for separating elements of a list or a map in case where it improves readability.

B. Infix operators

One of the most annoying aspects of S-expressions is the absence of infix operators. Prefix notation for arithmetic operators is very unusual and hard to read. So it is natural that there was a lot of attempts to introduce infix operators into Lisp.

Lets look at one of the implementations called Infpre [11]. It provides functions `infix->prefix` and `prefix->infix`. They allow to convert expression written in infix form to prefix form and vice versa. As the second parameter these functions take a list of valid operators, where list order gives precedence. For example:

```
% (infix->prefix '(1 + 2 * 3 + 4) '(+ *))
(+ 1 (* 2 3) 4)
```

The library also provides `math` macro allowing to define function with infix operators and `!!` macro, which applies `infix->prefix` with standard set of mathematical operators.

V. ALTERNATIVE SYNTAXES

Some Lisp dialects, like Logo or Dylan, went even further and abandoned S-expressions at all. They introduced new syntaxes trying to keep some advantages of S-expressions. There are also efforts to create alternative syntax for Common Lisp or Scheme which are closer to S-expressions, but try to solve some of their problems. Examples of such syntaxes are I-expressions, sweet-expressions or G-expressions.

A. Logo

Logo programming language [12], [13] was created in 1967. It was designed for teaching programming. Logo is based on Lisp, but uses different syntax, so it is some times called "Lisp without parentheses".

Logo syntax is based on the fact, that functions take fixed number of arguments. This allows to leave out parentheses in expressions. For example `print sum product 5 6 12` is equivalent to `print (sum (product 5 6) 12)`. Using explicit parentheses it is possible to pass different number of arguments to some build-in functions like `print` or `sum`. Logo also provides infix notation for basic arithmetic operators.

Logo has two basic data types: words and lists. Numbers are just special case of words. Words are quoted using quotation mark, e.g. `"hello`. Unquoted words are considered function names and are evaluated accordingly. List literals are written in square brackets and their contents are automatically quoted, e.g. `[hello 42 3.14]`. This feature is used in Logo control structures. Expressions that need to be evaluated conditionally or repeatedly are enclosed in a list. For example *if* command may look like this:

```
if 5 > 2 [print "Hello]
```

If and other control structures does not use special syntax in Logo. Expressions they take as their arguments are not evaluated before function call because they are enclosed in brackets.

A function that does not follow normal Logo syntax rules is *to*, which is used to define new functions. Its usage is shown in the example below.

```
to hello :name
  print sentence "Hello, :name
end
```

In the example you can also see another syntactical element — variable reference using colon (`:name`). It is actually a shortcut for call to *things* function (`things "name`).

Logo also supports macros, which are similar to functions, but return a list of Logo instructions. These instructions are then evaluated in the context of macro's caller. Syntax of macro definition is similar to function definition, but uses `.macro` keyword instead of `to`. Macro arguments are evaluated by the same rules as function arguments.

B. Dylan

Dylan programming language was developed in early 1990s at Apple Computer. It is object-oriented dynamic language designed for use in application and systems programming [14].

Dylan derives from Common Lisp and Scheme and early versions used S-expressions based syntax. However, during language development it was replaced with ALGOL-like syntax (see Fig. 3). In spite of this, Dylan still provides powerful macro system similarly to other Lisp dialects.

Dylan syntax for data literals is derived from Lisp. Boolean values are represented in the same way as in Scheme. Pairs and lists are prefixed with # and use commas to separate items: `#(1 . 2)`, `#(1, 2, 3)`. There are also vectors, written in square brackets: `#[1, 2, 3]`. Similarly to Scheme, identifiers in Dylan can contain special characters like "-" or "?". Dylan also supports symbols as separate data type, but they are not used to represent functions and variables. They have two syntactic forms: `foo:` and `#"foo"`.

```
define function factorial(n :: <integer>)
  => (n! :: <integer>)
    if (n = 0)
     1
    else
     n * factorial(n - 1)
    end
end method;
```

Figure 3. Definition of factorial function in Dylan.

```
a)                              b)
  let                             let
    group                           group
      foo                             foo (+ 1 2)
        + 1 2                         bar (+ 3 4)
      bar                           + foo bar
        + 3 4
      + foo bar

c)
    (let
      ((foo (+ 1 2))
       (bar (+ 3 4)))
      (+ foo bar))
```

Figure 4. Example of I-expressions syntax: a) I-expressions, b) combination of I-expressions and S-expressions, c) only S-expression.

Syntax for expressions and statements is similar to convenient languages like Pascal. For example, function calls have form `f(x, y)`, infix operators are supported, etc.

Dylan supports syntactic macros, which are expanded at compile-time. Macros are defined using rewrite rules, where the left-hand side of a rule is a pattern that matches a fragment of code and the right-hand side of a rule is a template which is substituted for the matched fragment [15]. Dylan provides three types of macros:

- *Definition macros* allow to extend available set of definitions. They can have two forms:
  - define *modifiers* DEFINE-WORD *body-fragment* end
  - define *modifiers* DEFINE-WORD *list-fragment*

  Where DEFINE-WORD is a name of a macro.
- *Statement macros* allow to create new types of statements. They have following form:

  BEGIN-WORD *body-fragment* end
- *Function macros* syntactically resemble function calls but are more flexible. They have following form:

  FUNCTION-WORD *(body-fragment)*

C. I-expressions

There are several attempts to create alternative syntaxes for Scheme and other Lisp dialects, that are normally based on S-expressions. One such attempt is called I-expressions. It is new syntax for Scheme, that uses indentation to group expressions. I-expressions are specified in SRFI 49 [16].

I-expressions have equal descriptive power as S-expressions and have no special cases for semantic constructs of the language. They are also mostly compatible with S-expressions, so it is possible to mix both syntaxes.

Using I-expressions, it is not needed to wrap expression in parentheses if it is written on separate line. If lines are indented, they are appended to the expression on previous line. Special keyword *group* is added to allow expressing list of lists (see Fig. 4).

```
define factorial(n)
  if {n <= 1}
    1
    {n * factorial{n - 1}}
```

Figure 5. Definition of factorial function using sweet-expressions.

```
define factorial (n)
  if (<= n 1)
    trace "base case"
    1
  else:
    trace "recursive case"
    * n (factorial (- n 1))
```

Figure 6. Definition of factorial function in Ginger.

D. sweet-expressions

David A. Wheeler created a collection of reader macros and filters for Scheme, Common Lisp and other Lisp dialects, that provide more readable syntax [17]. This collection includes:

- *Curly infix* interprets expressions enclosed in curly braces as infix expressions. For example {2 * 3 * 4} is translated to (* 2 3 4), and {2 + {3 * 4}} to (+ 2 (* 3 4)). There are no precedence rules, so subexpressions need to be enclosed in parentheses.
- *Modern-expressions* include curly infix and add special meaning for prefixed parentheses, brackets and braces. For example f(1 2) is translated to (f 1 2). So this allows to use function call notation used in mathematics and most other programming languages.
- *Sweet-expressions* include modern-expressions and add special meaning for indentation similar to I-expressions (see Fig. 5).

E. G-expressions

Another attempt to create indentation based syntax for Lisp based language is called G-expressions. It was developed for a new language called Ginger [18].

G-expressions are similar to I-expressions or sweet-expressions in that they use indentation for grouping. However, the fact that they were developed for new language allowed to take different approach in some aspects. G-expressions are not just different notation for S-expressions, but they have also different internal representation.

G-expressions threat indented blocks quite differently compared to I-expressions and sweet-expressions. In former, each line of indented block is separate argument for function call before indentation. In G-expressions block as a whole is considered single argument of parent expression, as if they were inside Scheme *begin* function. This provides better support for imperative style of programming.

To allow function to take several blocks as arguments, labeled line continuations are used. They are introduced by a symbol with colon (:) at the end. In the example code on Fig. 6 you can see how labeled continuation is used in *if-else* statement. It also shows, that code blocks can be used without explicitly calling *begin* function.

To allow expression to span several lines, anonymous line continuations introduced by double dot (..) are provided.

VI. Conclusion

This paper discussed syntax of Lisp family of programming languages and several it alternatives used by Lisp dialects.

The goal for development of alternative Lisp syntax is to make it more convenient and easy to read while preserving its flexibility. Balancing these aspects is a difficult task.

There are some conclusions, that can be made:

- To make code clear, program structure can be defined without parentheses using line breaks and indentation.
- It is useful to support more convenient infix notation for operators (especially for arithmetic operators).
- To allow extensibility of the language, its syntax should not define notation for concrete statements and operators. Instead of this, it should provide generic shapes for language elements.

This knowledge can be used for development of generic language suitable for hosting domain-specific languages. This language would provide generic syntax and infrastructure for domain-specific languages in the same way as XML supports development of data formats and markup languages. Prototype of such generic language was presented in [19].

Future work include design of flexible and readable syntax for generic language, based on experience of Lisp community, and development of associated infrastructure.

References

[1] A. C. Kay, "The reactive engine," Ph.D. dissertation, 1969.

[2] G. L. Steele, "Growing a language," *Higher-Order and Symbolic Computation*, vol. 12, pp. 221–236, 1999.

[3] H. Abelson and G. J. Sussman, "Lisp: A language for stratified design," Cambridge, MA, USA, Tech. Rep., 1987.

[4] E. S. Raymond, *The new hacker's dictionary (3rd ed.).* Cambridge, MA, USA: MIT Press, 1996.

[5] J. McCarthy, "History of programming languages I," R. L. Wexelblat, Ed. New York, NY, USA: ACM, 1981, ch. History of LISP, pp. 173–185.

[6] ——, "Recursive functions of symbolic expressions and their computation by machine, Part I," *Commun. ACM*, vol. 3, pp. 184–195, April 1960.

[7] ——, *LISP 1.5 Programmer's Manual.* The MIT Press, 1962.

[8] G. L. Steele, Jr., *Common LISP: the language (2nd ed.).* Newton, MA, USA: Digital Press, 1990.

[9] M. Sperber, R. K. Dybvig, M. Flatt, A. van Straaten, R. Findler, and J. Matthews, *Revised [6] Report on the Algorithmic Language Scheme*, 1st ed. New York, NY, USA: Cambridge University Press, 2010.

[10] R. Hickey. Clojure. [Online]. Available: http://clojure.org/

[11] J. I. Vestgården. Infpre - a common lisp infix utility. [Online]. Available: http://folk.uio.no/jornv/infpre/infpre.html

[12] B. Harvey, *Computer Science Logo Style, Vol. 1: Symbolic Computing*. MIT Press, 1997.

[13] ——, *Berkley Logo User Manual.* University of California Berkley, 1993.

[14] A. Shalit, *The Dylan reference manual: the definitive guide to the new object-oriented dynamic language.* Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc., 1996.

[15] J. Bachrach and K. Playford. (1999) D-expressions: Lisp power, dylan style. [Online]. Available: http://people.csail.mit.edu/jrb/Projects/dexprs.pdf

[16] E. Möller. Srfi 49: Indentation-sensitive syntax. [Online]. Available: http://srfi.schemers.org/srfi-49/srfi-49.html

[17] D. A. Wheeler. Readable s-expressions and sweet-expressions home page: Infix and fewer parentheses in lisp-like languages. [Online]. Available: http://www.dwheeler.com/readable/

[18] J. D. Palmer, "Ginger: implementing a new lisp family syntax," in *ACM-SE 47: Proceedings of the 47th Annual Southeast Regional Conference*. New York, NY, USA: ACM, 2009, pp. 1–6.

[19] S. Chodarev and J. Kollár, "XHL: DSL development framework based on fixed syntax," in *CSE 2010: proceedings of International Scientific Conference on Computer Science and Engineering*. Košice: Elfa, 2010, pp. 141–148.

Low-cost high throughput computing in cluster or grid environment using GPU

[1]*Marián MIŽIK*, [2]*Marek DOMITER*

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]marian.mizik@tuke.sk, [2]marek.domiter@tuke.sk

Abstract—**Parallel task computing on clusters and grid architectures is a partial solution of a problem, when throughput, capacity and performance reach its known architectural and construction limits. Dividing tasks to a partial subtasks, if it is possible, is lowering the amount of time needed for result computation and is also distributing the main task to be computed in the same time. The result and is this proposal of architecture of HTC Cluster based on parallel use of standard mainstream GPU cards to achieve the performance of a high tech cluster using software of hardware task redistribution.**

Keywords—**Cluster, HTC, GPU, GRID, SAN, NAS, MMP, SMP**

I. INTRODUCTION

Amost all scientific and research programs and their computing quality and quantity of research is strongly linked to computing throughput. Put differently, most scientists are interested in how many floating point operations per week or per month they can get from their computing environment rather than the number of such operations the environment can provide them at short time. The key to high throughput computing (HTC) is to efficiently manage and use all available computing resources. Computing needs are in these days represented by the processor and memory-intensive characteristics. High efficiency in this environment plays a major role. The main task, that typical HTC (High Troughput Computing) environment faces is how to maximize the amount of resources available for their customers. Distributed ownership of computing resources is a major barrier that must be overcome in order to broaden the portfolio of resources. Recent trends in price/performance ratio of computer hardware have placed the control of powerful computing resources into the hands of individual persons and small groups.

II. HIGH THROUGHPUT COMPUTING

A. Distributed tasks processing

If it is possible to divide the computational problem to many smaller subproblems, which can be solved in parallel, then distributing subtasks to different computer systems can accelerate the calculation. Assuming that these tasks are "embarassingly parallel" (maximally parallelizable), Amdahl's Law defines the highest possible acceleration achievable. Acceleration is given by the equation (1) (P is the proportion that is parallelizable; S is the expected acceleration of the task). This solution is called distributed computing, and this term refers to any system in which a set of inter connected computers is used to solve a problem. There are different types of distributed computing, grid computing such as cluster computing, peer-to-peer computing, Internet computing, meta, and so ones.

$$\frac{1}{(1\text{-}P)+\dfrac{P}{S}} \quad (1)$$

B. GPGPU

GPGPU stands for *General-Purpose computation on Graphics Processing Units*. Graphics Processing Units (GPUs) are high-performance many-core processors that can be used to accelerate a wide range of applications.

This architecture is very close to MPP (Massive Parallel Processing) and for its operation uses the graphics processor. Current graphics cards contain unified units (shaders) instead of a few specialized groups. These units are designed to carry small group of floating point calculations oriented on graphics operations. [1] The big advantage is that small group of calculations can be executed much faster than the CPU, which is more appropriate to carry out general calculations. Currently this technology is available for graphics cards manufactured by AMD and nVidia. [2]

From architectural point, CPU has much more control logic (one control unit per core) on the other hand, GPU has one crontrol unit for several FPU and ALU in parallel (a SIMD architecture) Traditional CPUs are oriented on several independent threads capable of high level of instruction parallelism and with support for block instructions (like sse, mmx and so on). SIMD architecture on GPU is great for alghorithms which need to execute same operation multiple times on different data. From these facts we can see that there are several important restrictions when choosing particular architecture. Together with growing complexity of instruction layer is more and more difficult to achieve parallelism on instruction layer (ILP). And at last but not least even when the algorithm is suitable for SIMD architecture, we can see performance degradation due to threads that diverges during the computation, because divergent threads should be compute in series.

GPU compared to CPU offer a lot of very fast cores that are specialized for floating point computation in simple precision

(simple precision - 32 bit), but in double precision (double precision - 64 bits) counting is slow, which is useless for scientific calculations. Manufacturers are aware of this problem and the latest GPUs are trying to eliminate this limitation. An example is the Tesla GPU c2050 nVidia known as Fermi. For scientific calculations and in terms of efficient use of resources would be advisable, to run more tasks on one card even when GPUs were not built to compute several tasks in parallel. Latest GPU cards already counts with this usage (Fermi).

III. CLUSTER ARCHITECTURE

A. Architecture choice

To decide what type of approach is chosen, the type of application that will run on a cluster takes effect. If the application is developed by cluster owner or source code can be modified, MPP should be used as a solution. This is a set of tools to help with the applications dividing for sub-units. If a situation arises where the application is not possible to perform interventions, it is preferable to use SMP or NUMA as in Fig. 1, because there is load distribution ensured by operating system. Main type of the tasks should be considererd during choosing the right architecture.
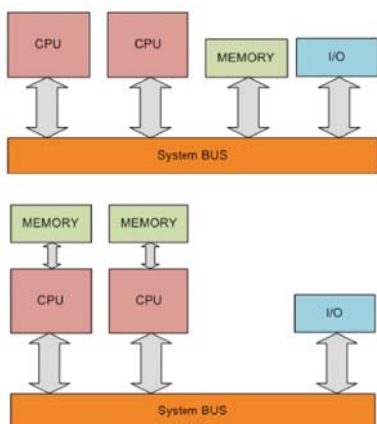
Fig. 1. SMP (upper) and NUMA architecture. SMP with shared memory and NUMA with the memory location relative to a processor.

B. Middleware

Middleware is computer software that connects software components or applications and creating a communication and sharing layer. The software consists of a set of services that allows multiple processes running on one or more machines to interact between each other. Having the proper cluster middleware can make the difference between 20% utilization of your system and 99% utilization. Furthermore, cluster middleware such as MPI or PVM centric or similar technologies are key enabling parallel processing, and other cluster middleware pieces are used to monitor hardware and automate software deployment and other tasks. [3]

C. Grid Middleware

Grid middleware is a specific software product, which enables the sharing of heterogeneous resources, and Virtual Organizations. It is installed and integrated into the existing infrastructure of the involved company or companies, and provides a special layer placed among the heterogeneous infrastructure and the specific user applications.

IV. HTC CLUSTER

As high throughput is required, one supercomputer is less efficient when processing huge amount of data and is necessary to process data on multiple cooperated computers - cluster nodes. HTC clusters - compared with other types (high availability clusters, load balancing clusters and others) are focusing on scalability of systems with large data throughput and arithmetic performance, which determines the requirements for the system architecture. Capacity of cluster can be increased by donating additional GPU cards installed on worker nodes, where jobs will be computed. These jobs must be already implemented in ready way for massive parallel processing.

Network Connections
Nodes interconnections are characterized by:

- Bandwidth
- Latency

Standardly inexpensive Ethernet technology can be used with communication over IP protocol. This solution is cheap and bandwidth can be extended using bonding – IEEE 802.3ad standard. In case of standard servers bandwidth 2Gbit/s is achieved.
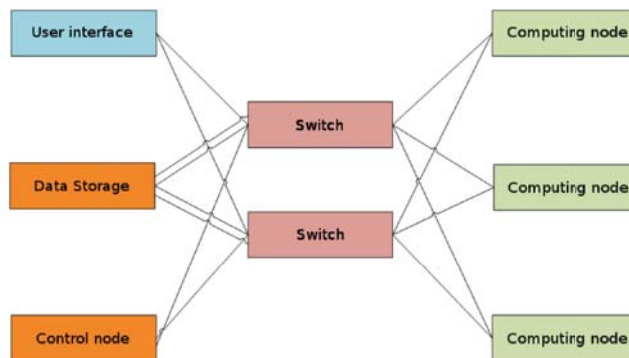
Fig. 2. Proposed HTC cluster architecture.

Data storage
Information in cluster can be stored centralized or decentralized – distributed. Centralized data storage is preferable in case of small cluster. Availability of data is ensured by *Redundant Array of Independent Disks(RAID)* on central storage node.

Storage is accessed by one of method:

- **storage area network** (SAN) – a dedicated storage network which provides access to consolidated, block level storage

- **network area storage** (NAS) – a single storage device that operate on data files

Workload management

The key to HTC is effective management and exploitation of all available computing resources. Since the computing needs of most scientists can be satisfied these days by commodity CPUs and memory, high efficiency is not playing a major role in a HTC environment. The main challenge a typical HTC environment faces is how to maximize the amount of resources accessible to its customers. Distributed ownership of computing resources is the major obstacle such an environment has to overcome in order to expand the pool of resources it can draw from. Recent trends in the cost/performance ratio of computer hardware have placed the control (ownership) over powerful computing resources in the hands of individuals and small groups. These distributed owners will be willing to include their resources in a HTC environment only after they are convinced that their needs will be addressed and their rights protected.

While providing functionality similar to that of a more traditional batch queuing system, novel architecture allows it to succeed in areas where traditional scheduling systems fail. Solution can be used to manage a cluster of dedicated compute nodes. In addition, unique mechanisms enable Condor to effectively harness wasted CPU power from otherwise idle desktop workstations. For instance, Condor can be configured to only use desktop machines where the keyboard and mouse are idle. Should Condor detect that a machine is no longer available (such as a key press detected), in many circumstances Condor is able to transparently produce a checkpoint and migrate a job to a different machine which would otherwise be idle. It does not require a shared file system across machines - if no shared file system is available, then it transfers the job's data files on behalf of the user, or Condor may be able to transparently redirect all the job's I/O requests back to the submit machine. Result then can be used to seamlessly combine all of an organization's computational power into one resource.

When users submit their serial or parallel jobs, workload manager places them into a queue, chooses when and where to run the jobs based upon a policy, carefully monitors their progress, and ultimately informs the user upon completion. As workload manager can be used condor. [4] On Fig. 2. daemons of condor are showed with dependencies. For HTC computing new universe need to be created for native GPU jobs support.
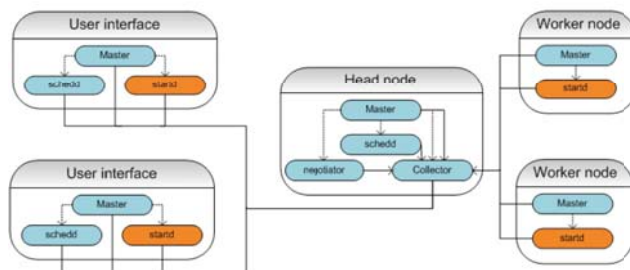
V. CONCLUSION

Cluster and grid computing abets solving for nonlinear processor power increasing problem in actually known hardware technologies. Dividing tasks into parallel subtasks (if possible) shorten computation time and give higher data throughput to processing jobs.

The result is high throughput counting that within long periods of time (month, year) will process petabytes of data at the perform of peta flop per second. Graphics card should contribute to the HTC, where the high data throughput and processing power makes them the most powerful commonly available equipment.

REFERENCES

[1] D. Luebke, "GPU Architecture: Implications & Trends." [Online] 2008. [Date: 11. 12 2010.] http://s08.idav.ucdavis.edu/luebke-nvidia-gpu-architecture.pdf.

[2] L. Vokorokos, J. Perháč, B. Madoš, "Distributed GPGPU – Implementation Possibilitie"s. Stará Lesná : Elfa s.r.o., 2008. ISBN 978-80-8086-089-9.

[3] LinuxHPC.org/Cluster Builder 1.3. [Online] Adaptive Computing . [Date: 20. 11 2010.] http://www.clusterbuilder.org/.

[4] Condor Project. [Online] UW-Madison Computer Sciences Department. [Date: 10. 12 2010.] http://www.cs.wisc.edu/condor/.

[5] L. Vokorokos, Parallel Computing Systems applied to Secure Architectures and Computer Networks. Košice : Elfa s.r.o., 2007. ISBN 978-80-8086-046-2.

[6] R. J. Hovland, Latency and Bandwidth Impact on GPU-system. 2008

[7] J. Perháč, "Design of GPGPU utilization in distributed computer system 7th PhD Student Conference and Scientific and Technical Competition of Students of Faculty of Electrical Engineering and Informatics Technical University of Košice, Košice, Slovakia, 23.5.2007, Košice, Slovakia, Elfa, 2007, 7., pp. 157-158

Fig. 3. Architecture of HTC clusters workload manager. User interfaces for jobs submitting, Head node and Worker nodes – represent cluster nodes.

Maintenance of Knowledge Based Software System

[1]Peter ŽÁRSKÝ

[1]Dept. of Computer Science and Informatics, FEI TU of Košice, Slovak Republic

[1]peter.zarsky@tuke.sk

Abstract—**This paper presents potential of concept methodology of knowledge based software system during the maintenance process. The concept briefly described in the second chapter is to ease the maintenance process and to help avoid clogging errors into source code by higher level of abstraction.**

Keywords—**software system architecture, knowledge based software systems, model driven maintenance.**

I. INTRODUCTION

Software engineering is an iterative process that comprises multiple stages, including modeling, design, implementation, deployment, and maintenance. Choosing an architecture design in this process is one of the most important things. Regardless of the variations in software engineering processes, software architecture provides the skeleton and constraints for software implementation. Rising complexity of software system declines its understandability and become less easy to handle it. Research in the present is focused to knowledge about the software system and their interconnection with code of application.

The methodology concept described in the next chapter was designed to allow Model Driven Maintenance (MDM). [1, 2] This process is useful aspect of knowledge-based software life cycle oriented to better usability of all analysis, design and implementation models in maintenance of systems. It tries to streamline and speed up the maintenance process of software system together with keeping system consistency with the help of knowledge acquired from software system's abstract models. [3]

Models are cornerstones of MDM, and their consistency with other artifacts of the system, especially with the code is crucial. This is the reason for the proposal of a modified system architecture, which supports a conjunctive preservation of the program code and its models.

II. STRUCTURE OF THE CONCEPT

Structure of this concept [4, 5] consists of four main parts. Set of software system components, that is the core of the software system, base of knowledge, transforming script and GUI (Graphical User Interface) based tool. This provides access to the base of the knowledge for the designer or user, as shown on Fig. 1.

A. Set of Software System Components

This part of software system is the logical set of all components. It also includes external services or resources, which are used by the software system. This set does not consist only of components that are executable, reusable parts of code as in component based architecture. Parts of this set are all data resources used by the software system, configuration files, external services etc. Each component has its specification stored in the base of knowledge. Component itself does not contain any information about other component that is in active or passive communication with it. This information is read from the base of knowledge every time it is needed. It is read from the knowledge base at start of the execution of the component or immediately before the knowledge is needed, that depends on the component's demands. This ensures that components behave according to actual model stored in the base of knowledge.

B. Base of Knowledge

Key part of the concept for interconnection of implementation and documentation of the software system is the base of knowledge. This part of software system stores knowledge that is used by both: components and transformation scripts for GUI based tools that enable to reengineer the model of that software system. The knowledge is stored in base and can be divided into groups. In the first group, there are stored specifications and notes of all components from set of components. This knowledge is used for accessing documentation of software system and during modification of software system (adding new components, checking integrity of software system). For this knowledge access time is not critical and for storing them can be used any DBMS (database management system). In the second group, there is a compact knowledge about the relationships and communication channels between all components from the set of components. Compared to the first group of knowledge, this group includes knowledge needed during execution of components therefore, access time is critical. The amount of data stored in this part of information system is not enormous, so they can be cashed in main memory for faster access.

C. Transformation scripts

Knowledge stored in the knowledge base need to be transformed to format, which can be viewed by GUI based tools. For this purpose there are transformation scripts, which transform knowledge form knowledge base into diagrams and store them file. These scripts are also responsible for transforming changes made in file into knowledge base.

For better compatibility with GUI based tools is model stored in XML Metadata Interchange (XMI) standard developed by Object Management Group (OMG). [6]
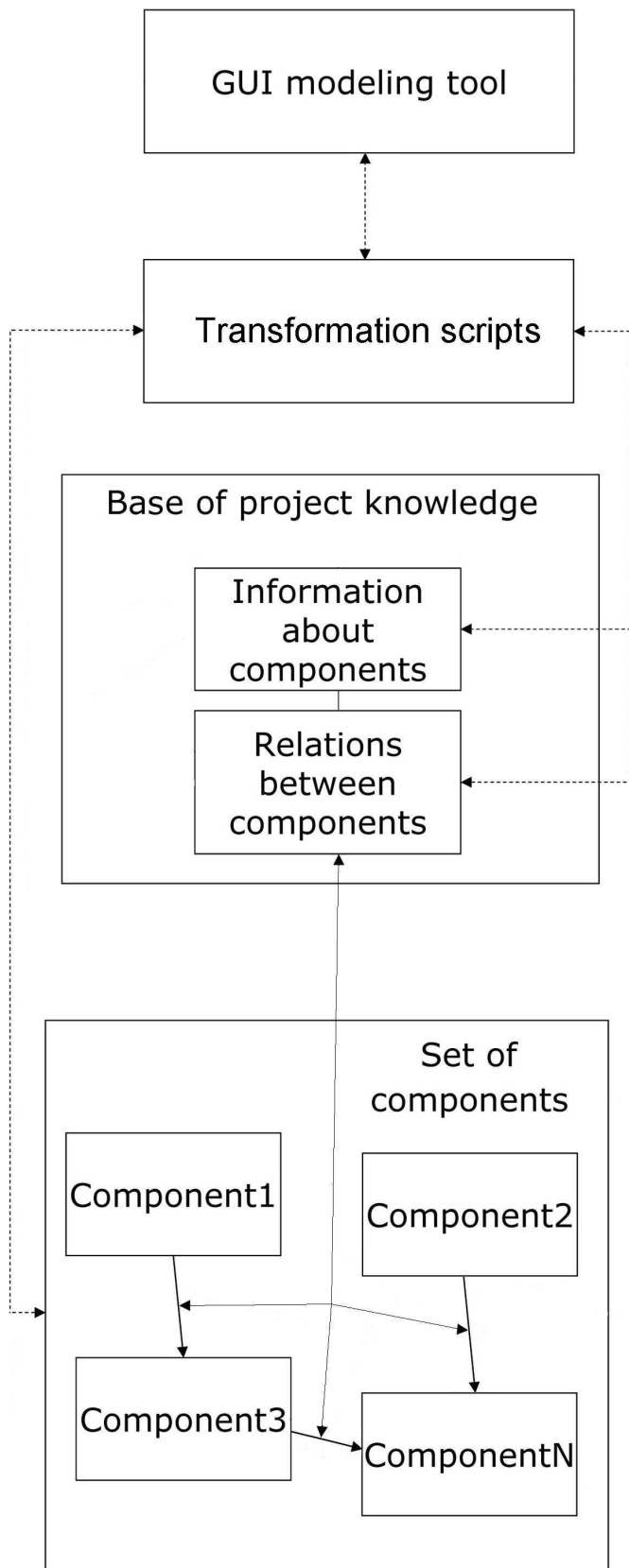
Fig. 1. Structure of the architecture concept.

D. GUI modeling tool

The last part of this concept architecture is tool with GUI that allows developer or user to view and change model of software system. This tool will view the models of software system transformed from base of knowledge by transformation scripts, so it will always correspond with the current version of the software system. This tool with transformation scripts should allow adding and removing of components from set of components, maintain relations between components and check integrity of the system. Integrity of the system can be corrupted in two ways. User can try to remove a component that has defined relation with another component. This can lead to situation, when component needed by another component is missing in the system. Another case of corrupting the system integrity is, if user will try to add new component into set of components without defining all relations of this component described in its interface specification. It would cause the same situation as previous case.

III. Conclusion

Maintaining and upgrading of evolving software system can be difficult and expensive process. Knowledge based systems and MDM approach eases this process and prevent some bugs that can be made during maintenance.

References

[1] Kunštár, J., Adamuščinová,I., Havlice,Z.: Model-Driven Life Cycle, CSE'2008 International Scientific Conference on Computer Science and Engineering, High Tatras – Stará Lesná, Slovakia, Semtember 24-26, 2008, ISBN 978-80-8086-092-9.

[2] Samuelis L., Havlice Z., Telepovská H., Kunštár J., Adamuščinová I., Pločica O., Varga M., Železník O., Révés M., Huňady M.: Software Processes Based on Knowledge, Computer Science and Technology Research Survey, Košice, Department of Computers and Informatics, FEII TU of Košice, 2008, 1, 3, pp. 11-18, ISBN 978-80-8086-100-1

[3] Havlice Z., Kunštár J., Adamuščínová I. and Pločica O.: Knowledge in Software Life Cycle, SAMI 2009 Proceedings, 7th International Symposium on Applied Machine Intelligence and Informatics, Herľany, Slovakia, 30.-31.1.2009, 2009, IEEE Catalog Number CFP0908E-CDR, ISBN 978-1-4244-3802-0

[4] Žárský P., Lakatoš M., Havlice Z.: The Role of the Project-Knowledge in IS Development, Journal of Computer Science and Control Systems, Academy of Romanian Scientists, University of Oradea, Faculty of Electrical Engineering and Information Technology, Vol. 2, Nr. 2, 2009, pp. 80-83, ISSN 1884-6043

[5] Žárský P., Lakatoš M., Havlice Z.: Integration of Knowledge about Software System and Used Design Patterns into Architecture of Software System, Proceedings of CSE 2010 International Scientific Conference on Computer Science and Engineering, Stará Ľubovňa, September 20-22, 2010, Košice, elfa, s.r.o., 2010, 1, pp. 247-251, ISBN 978-80-8086-164-3

[6] Object Management Group: XMI specification, accessible on the internet http://www.omg.org/spec/XMI/Current

Neural Networks and Financial Time Series Forecasting

[1]*Pavol MALIŇÁK*, [2]*Marek LAPKO*, [3]*Peter SINČÁK*

[1,2,3]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

{[1]pavol.malinak, [2]marek.lapko, [3]peter.sincak}@tuke.sk

Abstract—In this paper, properties of financial time series with impact on both design of neural networks and accuracy of forecasts are investigated. Simple tests for assessing time series predictability are described. Results from our experiments indicate that artificial neural networks are able to achieve a remarkable forecasting accuracy on the financial data.

Keywords—forecasting, predictability, rescaled range analysis, time series

I. INTRODUCTION

Time series forecasting is recognized as especially challenging and interesting area with applications in various fields of science, engineering, medicine, finance and others. The fact that markets clearly reflect also diverse social, economic and political affairs causes that financial time series (especially from mature stocks) often behave like a random walk (RW). This makes them difficult to model and further causes that financial time series forecasting is generally considered as "a very specific task" [1], [2], [3], [4]. As stated in [5], though the issue of the nature of financial data is still an open question, they are often regarded to be nonlinear stochastic, chaotic or a combination of both (cf. [6], [7]).

In this paper, we let aside the contradictory views of Efficient Market Hypothesis[1] and adopt the technical analysis approach. Given the current and past values of a sequence of observations, the goal is to model the history in order to predict (estimate) the future data values in the selected horizon. Since this is inherently a function approximation task, the use of neural networks, with respect to their ability to model complex nonlinear relationships, is regarded as a natural solution.

II. NEURAL NETWORKS

As we have already stated above, neural networks are known for their ability to model complex nonlinear relationships. Unlike traditional stochastic models, which do not perform satisfactorily on the financial time series [8], [5], neural networks do not rely on the assumption of linearity of the data generating process [9], [10]. Thus, being nonparametric universal function approximators, they are able to approximate any reasonable function with arbitrary accuracy. As such, they have been widely used in financial applications, often clearly outperforming other statistical and machine learning techniques [11].

[1]EMH states that (average) excess profits are just a result of luck – there is no difference between uninformed investor and professional. Contrary, market participants believe that there are strategies (from technical or fundamental analysis) which can be used to beat the market.

Handling of time series sequences requires some kind of memory to be present in the neural network. Incorporation of temporal information into the network can be carried out either explicitly (time-delay structures) or implicitly (using the spatial structure of the net). For these purposes time-delay (TDNNs), FIR, recurrent (RCNNs) and negatively correlated (NCNNs) neural networks are often considered in experiments. During the training process, standard backpropagation, Back-Propagation Through Time (BPTT), Real Time Recurrent Learning (RTRL) or Temporal Back-propagation are often used. In order to achieve reliable prediction abilities, incorporation of "local information" into network objective (error) function can also be encountered, e. g. in Violation-Guided Backpropagation (VGBP) algorithm. Second order learning techniques (e.g. Levenberg-Marquardt) are also widely used.

III. PREDICTABILITY ANALYSIS

An important issue which should be addressed before any forecasting attempt (during the preprocessing step) is that of predictability of time series at hand [7]. If time series is a pure random walk process (i.e. tomorrow's value is actual (today's) value plus disturbance), all predictions are useless. In such a case any algorithm will produce 50% hit rate in the long run (see e.g. [3]). In our case, the predictability has been assessed by use of rescaled range (R/S) analysis [12], [13], [4], [14], runs test [15] and the turning point test [16].

A. Rescaled Range Analysis

According to the R/S analysis, the memory characteristics of time series can be described by a statistical relationship

$$(R/S)_\tau \sim C\tau^H \tag{1}$$

as τ increases, where $H \in \langle 0; 1 \rangle$ is so-called Hurst Exponent, C is a constant, R_τ and S_τ are average range and standard deviation over a sub-period of length τ, respecitvely. To estimate the Hurst parameter, simple OLS regression is run over different values of sub-interval τ, i.e.

$$\log (R/S)_\tau = log(C) + H \log(\tau) \tag{2}$$

Based on the value of H index, time series can be classified as persistent ($0.5 < H < 1$), antipersistent ($0 < H < 0.5$) and random ($H = 0.5$). However, it should be stressed, as pointed out in [8] and [12], that the value of $H = 0.5$ does not provide us with the evidence that a particular series is the RW – it just tells us that there is no evidence for a memory effect.

B. Turning Point Test

This test of randomness is based on inspection of the number of local maxima and minima (so-called turning points) present in the time series. If the data really is random, one can find out the expected number of turning points and compare it with observed value T for a particular time series $\{x_1, x_2, \ldots, x_N\}$ of N observations.

In the case of a purely random process, the triple x_{t-1}, x_t, x_{t+1} is equally likely to occur in any of the six possible orders with the probability of turning point at time t equal to $2/3$. This leads to the expected number of turning points $\mu_T = \frac{2}{3}(N-2)$ and variance $\sigma_T^2 = \frac{1}{90}(16N - 29)$. Since the number T of turning points should be approximately normally distributed, the test statistic can be defined as

$$\frac{|T - \mu_T|}{\sigma_T} < z_{1-\alpha/2} \quad (3)$$

If T violates the condition (3), the null hypothesis of randomness is rejected at the level α.

C. Wald-Wolfowitz Runs Test

The runs test is a simple nonparametric technique to test the randomness of a time series by inspecting the frequency of some repetitive patterns in the series. Similarly to the turning point test, null hypothesis states that the elements of a time series are mutually independent.

Particular series $\{x_1, x_2, \ldots, x_N\}$ of N observations can be associated with the sequence of N symbols '+' or '−' derived by writing $+$ when x_n is greater than the median value, and $−$ otherwise. A run is a sequence of consecutive $+$ or $−$ symbols. Further, let R be the number of runs, n_1 be the number of occurrences of $+$, n_2 be the number of occurrences of $−$ and $n_1 + n_2 = N$. Under the null hypothesis of independence, it can be shown that $\mu_R = \frac{2n_1 n_2}{n_1 + n_2} + 1$ and $\sigma_R^2 = \frac{2n_1 n_2 (2n_1 n_2 - N)}{N^2(N-1)} = \frac{(\mu_R - 1)(\mu_R - 2)}{N-1}$ and the distribution of R should by standard normal. Thus, if R violates the condition

$$\frac{|R - \mu_R|}{\sigma_R} < z_{1-\alpha/2}, \quad (4)$$

the null hypothesis of randomness is rejected at the level α.

IV. EXPERIMENTS

During our research, certain private company provided us with 46 "unknown" financial time series. Our goal is to reliably predict the short-term profit opportunities of a given stock by use of neural networks. The profit opportunity is understood as certain (e.g. 0.5%, 2%) change in the price within particular time period. Thus, we are much more interested in change-in-direction forecasting, then in the price forecasting.

This aim motivated us to a simple modification of TDNN error function: In each iteration, the resulting squared error of the network has been multiplied by a certain factor, κ, in order to penalize the forecasts of profit opportunities that are incorrect in the direction of price movements.

Number of inputs to be fed into the network, as well as their separation in time, has been found on the basis of nonlinear analysis of the data [17]. Separation (time delay) has been estimated in two ways – as a lag where autocorrelation function shows first zero correlation, and as the lag where first minimum of average mutual information function occurs.
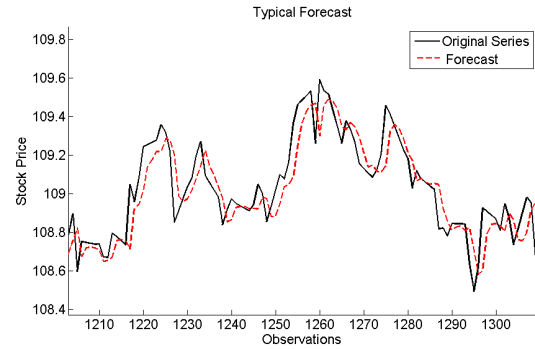
Fig. 1.   An example of undesirable forecast – network systematically "predicts" the past values of the series. This behavior is typical e.g. for TDNN prediction.
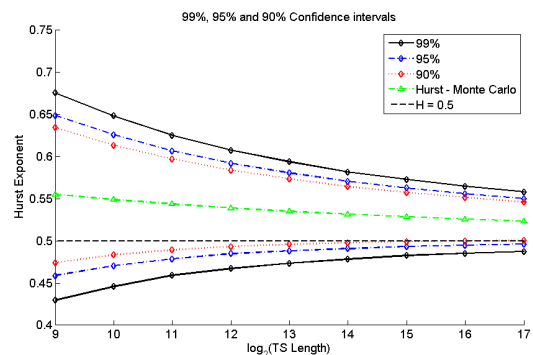
Fig. 2.   Confidence intervals for Hurst coefficient based on our Monte Caro simulations. Here, time series of lengths from 2^9 to 2^{17} have been investigated. The minimum scale is set to $\tau_{\min} = 2^3$ observations and maximum scale was set to $\tau_{\max} = \text{floor}(\log_2(N))$ observations, where N is the length of the series. For each time series' length, 10^5 random Gaussian $N(0, 1)$ series were generated.
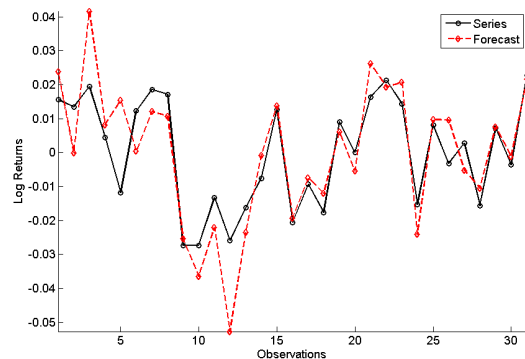
Fig. 3.   Forecasting behavior obtained by our modification of neural network criterial (error) function.

Number of inputs (embedding) has been set by use of false nearest neighbors method.

Based on the results from extensive experiments, neural networks topologies with one hidden layer have been adopted. Interestingly, larger topologies have demonstrated a negative impact on the quality of forecasts. Generally, many of our findings are consistent with those previously reported by Walczak [11].

V. CONCLUSIONS

Results of our experiments indicate that (under certain circumstances) neural networks are able to achieve a remarkable

TABLE I
PREDICTABILITY ANALYSIS

Time Series	Hurst	Wald-Wolfowitz test p-value	Turning point test p-value
1	0.4679	≈ 0	0.0302
11	0.4480	0.0450	0.2110
12	0.4733	0.0280	0.0875
23	0.5370	0.0001	0.0030
24	0.5251	0.0071	0.0043
32	0.4113	0.0282	0.0401
41	0.4796	0.0001	0.0087

Predictability analysis of the logarithmic returns for particular financial series.

forecasting accuracy on the financial data – up to 56%-57%. While promising, they are subjected to the careful statistical evaluation. The results have also confirmed that the "degree" of randomness (detected via various predictability tests) of a particular time series influences the forecast accuracy. Time series that have successfully passed through the predictability tests could be predicted with considerably higher accuracy.

REFERENCES

[1] A. Lendasse, E. de Bodt, V. Wertz, and M. Verleysen, "Non-linear financial time series forecasting – Application to the Bel 20 stock market index," *European Journal of Economic and Social Systems*, vol. 14, no. 1, pp. 81–91, 2000. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.11.9681

[2] J. Yao, C. L. Tan, and H. L. Poh, "Neural Networks for Technical Analysis: A Study on KLCI," *International Journal of Theoretical and Applied Finance*, vol. 2, no. 2, pp. 221–241, 1999. [Online]. Available: http://www2.cs.uregina.ca/~jtyao/Papers/klci_ijtaf.pdf

[3] T. Hellström and K. Holström, "Predicting the Stock Market," Center of Mathematical Modeling, Department of Mathematics and Physics, Mlardalen University, Technical Report IMa-TOM-1997-07, August 1998. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.57.4327

[4] B. Qian and K. Rasheed, "Hurst Exponent and Financial Market Predictability," in *Proceedings of The 2nd IASTED International Conference on Financial Engineering and Applications, Cambridge, MA, USA*, November 2004, pp. 203–209. [Online]. Available: http://qianbo.myweb.uga.edu/research/Hurst.pdf

[5] F. Strozzi, E. Gutirrez, C. Noè, T. Rossi, M. Serati, and J. M. Zaldvar, "Application Of Non-Linear Time Series Analysis Techniques To The Nordic Spot Electricity Market Data," LIUC Universitá Carlo Cattaneo, Report, October 2007. [Online]. Available: http://www.manmadenet.eu/publications/Report_Nordpool_LIUC_final.doc

[6] E. E. Peters, *Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility*. John Wiley & Sons, 1991. [Online]. Available: http://books.google.com/books?id=NszGthtCiU0C&dq=%22chaos+and+order+in+the+capital+markets%22&source=gbs_navlinks_s

[7] M. L. Qian, "Neural Network Learning for Time-Series Predictions Using Constrained Formulations," Ph.D. dissertation, Department of Computer Science, University of Illinois at Urbana-Champaign, May 2005. [Online]. Available: http://manip.crhc.uiuc.edu/Wah/papers/TP17/TP17.pdf

[8] J. Yao and C. L. Tan, "A case study on using neural networks to perform technical forecasting of forex," *Neurocomputing*, vol. 34, no. 1-4, pp. 79–98, September 2000.

[9] T. A. Wong, "Testing the Weak Form Efficiency of Shanghai Stock Exchange with Artificial Neural Networks," Master's Thesis, University of Nottingham, 2007. [Online]. Available: http://edissertations.nottingham.ac.uk/1512/1/07MSclixtcaw.pdf

[10] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, March 1998.

[11] S. Walczak, "An Empirical Analysis of Data Requirements for Financial Forecasting with Neural Networks ," *Journal of Management Information Systems*, vol. 17, no. 4, pp. 203–222, 2001. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.23.6904

[12] L. Kritoufek, "Distinguishing between short and long range dependence: Finite sample properties of rescaled range and modified rescaled range," Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, MPRA Paper 16424, July 2009. [Online]. Available: http://mpra.ub.uni-muenchen.de/16424/2/MPRA_paper_16424.pdf

[13] ——, "R/S Analysis and DFA: Finite Sample Properties and Confidence Intervals," Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, MPRA Paper 16446, July 2009. [Online]. Available: http://mpra.ub.uni-muenchen.de/16446/1/MPRA_paper_16446.pdf

[14] H.-w. Sang, T. Ma, and S.-z. Wang, "Hurst Exponent Analysis of Financial Time Series," *Journal of Shanghai University (English Edition)*, vol. 5, no. 4, pp. 269–272, December 2001. [Online]. Available: http://www.springerlink.com/content/r1520312v9667410/

[15] C. Chatfield, *The Analysis of Time Series: An Introduction, Fifth Edition*. Chapman & Hall/CRC Press, 1996.

[16] P. Castiglioni and M. Di Rienzo, "How to check steady-state condition from cardiovascular time series," *Physiological measurement*, vol. 25, no. 4, pp. 985–996, August 2004.

[17] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, "The analysis of observed chaotic data in physical systems," *Reviews of Modern Physics*, vol. 65, no. 4, pp. 1331–1392, October 1993.

Object calculus for formal description of software object

[1]*Marián JENČIK*

[1]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]marian.jencik@student.tuke.sk

Abstract—This paper introduces object calculus as advancement of λ- calculus by object primitives. This calculus supports *object extension* that is the ability of software object behaviour modification by adding new methods and inheriting the existing ones. Behaviour of the objects can be modified by method redefining, too. This ability is named *object modification*. However object calculus lacks one of the most important features of object-oriented programming, namely full subtyping mechanism, because we cannot use subtyping when method adding or method redefinig is allowed, there is the restricted subtyping mechanism with two types : *width subtyping* - is abled only if method redefining operation is denied, *depth subtyping* - is abled only if method adding operation is denied.

Keywords—object calculus, protype, suptyping

I. INTRODUCTION

Object calculus [1], [2] is based on λ- calculus [3], [4], formulated by Alonzo Church in 1930 and is advanced by object primitives. Operations allowed on objects are :

- operation *method adding* - for new methods defining;
- operation *method redefining* - for methods redefining;
- operation *method invocation* - for sending message m to the object e.

Software objects are built up from an empty object by adding new methods or redefining existing ones. The calculus supports a inheritance mechanism, a straightforward method specialization, and dynamic invocation of the methods.

Method are specialized by static type system which introduces an elegant form of method specialization whereby the type of inherited methods may be specialized to the type of the host objects.

II. UNTYPED OBJECT CALCULUS

Starting point of object calculus description is for its simplicity untyped object calculus, namely untyped λ- calculus extended by four object-oriented forms:

$\langle\rangle$	empty object;
$e \Leftarrow m$	sending message m to the object e;
$\langle e_1 \longleftarrow\!\!\!+ m = e_2 \rangle$	object e_1 extension by method m with body e_2;
$\langle e_1 \leftarrow m = e_2 \rangle$	method m body e_1 redefine (override) by body e_2.

Syntax of expresions e, thus extended untyped λ- calculus, are defined by production rule as follows :

$$e ::= \quad x \mid c \mid \lambda x.e \mid e_1 e_2 \mid \langle\rangle \mid e{\Leftarrow}m \mid \langle e_1{\leftarrow}m{=}e_2\rangle \mid$$
$$\langle e_1 \longleftarrow\!\!\!+ m{=}e_2\rangle.$$

In this grammar is x term variable, c is a constant, $\lambda x.e$ is function definition, $e_1 e_2$ is function application and m is the name of the method.

Example 2.1 (Methods inheritance between objects): Let's have object $point$ with methods x and $move_x$[1] :

$$point \overset{def}{=} \quad \langle\ x{=}\lambda self.3,$$
$$move_x{=}\lambda self.\lambda feed_x.\langle self{\leftarrow}x{=}\lambda s.(self{\Leftarrow}x){+}feed_x\rangle$$
$$\rangle.$$

where $feed_x$ is value of x-coordinates incrementation.

We can create object $pointA$ by adding method *color* and inheriting methods x and $move_x$ from object *point* :

$$pointA \overset{def}{=} \langle point \longleftarrow\!\!\!+ color{=}\lambda self.red\rangle.$$

If we send message $move_2$ to the object $pointA$, we can calculate result as follows :

$$
\begin{aligned}
pointA{\Leftarrow}move_2 &= (\lambda self.\lambda move_x\langle...\rangle)pointA_2 \\
&= \langle pointA{\leftarrow}x{=}\lambda s(pointA{\Leftarrow}x){+}2\rangle \\
&= ... \\
&= \langle pointA{\leftarrow}x{=}\lambda self.5\rangle.
\end{aligned}
$$

with the conclusion :

$$
\begin{aligned}
pointA{\Leftarrow}move_2 = \quad &\langle \\
x= \quad &\lambda self.5, \\
move_2= \quad &move_x{=}\lambda self.\lambda feed_x.\langle...\rangle, \\
color= \quad &\lambda self.red \\
&\rangle.
\end{aligned}
$$

Notation 2.1: By calculation it was used object equality rule :

$$\langle\langle m_1{=}e_1,...,m_k{=}e_k\rangle{\leftarrow}m_i{=}e_i\rangle{=}\langle m_1{=}e_1,...,m_i{=}e_i,...,m_k{=}e_k\rangle$$

and evaluation rules :

(β - reduction)	$(\lambda x.e_1)e_2$	$\overset{eval}{\to}$	$[e_2/x]e_1$,
($\Leftarrow$)	$\langle e_1{\leftarrow}m{=}e_2\rangle{\Leftarrow}m$	$\overset{eval}{\to}$	$e_2\langle e_1{\leftarrow}m{=}e_2\rangle.$

$\square$

A very important feature of this calculation is that the *color* of the resulting object $pointA$ is the same as the original. While $move_x$ method was defined for objects that have x-coordinates, body of the method allows calculation when the

---

[1]Special symbol self allows to introduce the special symbol self of object oriented languages directly to lambda abstraction.

method is inherited by more complicated object with method adding.

□

In many cases it is useful to redefine inherited methods to achieve desired behavior. This can be achieved in the λ-calculus by inheriting methods and redefining. For example, if we would like to change the color of the object in our example 2.1 to darker by object movement, we could define object $pointA$ from object $point$, as we can see above. Obtained object $pointA$ will contain orginal $move_x$ method. To obtain desired behaviour, $move_x$ method would be able to redefine by *color* method.

III. STATIC TYPE SYSTEM FOR OBJECT CALCULUS

Type of the object we can define according to [1], [5],[6] by type expression in the form. :

$$\textbf{pro } t.\langle\!\langle m_1 : \varrho_1, \ldots, m_k : \varrho_k\rangle\!\rangle$$
or in the reduction form
$$\textbf{pro } t.R, \quad \text{where} \quad R = \langle\!\langle m_1 : \varrho_1, \ldots, m_k : \varrho_k\rangle\!\rangle$$

and we name it **pro**type, where t is *type variable*. Thus defined type expression describes all objects, which contain method, respectively methods with the name m_i, that the result of message sending to the object ($e \Leftarrow m_i$) is value of ϱ_i type, for $1 \leq i \leq k$.

Type variable t take the values in the range of $\varrho_1, \ldots, \varrho_k$ and refers to the object itself.

The expression $R = \langle\!\langle m_1 : \varrho_1, \ldots, m_k : \varrho_k\rangle\!\rangle$ we name *row*. *Row* contains method names and their associated types, does not contain method bodies themselves.

Very important feature thus defined type system is that if the object is extended by method adding, type of each method does not change.

Protype of the object from example 2.1 where we had extended object $point$ with move function $move_x$ by *color* method , to obtain $pointA$ object we can write as follows :

$$\textbf{pro } t.\langle\!\langle x{:}int, move_x{:}int{\rightarrow}t, color{:}color\rangle\!\rangle$$

Significant change, that we obtain by defining the type system is that although type of the $move_x$ method is $int \rightarrow t$, variable t does not refer to the type of the object $point$, but to the object $pointA$ type. This feature is called method specialization and allows us to change the type of the method, if the method is inherited.

Set of all types contains type variables, function types and **pro**types, which we can defined as follows :

- **Types**

$$\varrho ::= \underbrace{t}_{\text{type variable}} \mid \underbrace{\varrho_1 \rightarrow \varrho_2}_{\text{function type}} \mid \underbrace{\textbf{pro } t.R}_{\textbf{pro}\text{type}}$$

- **Rows**

$$R ::= \varrho \mid \langle\!\langle\rangle\!\rangle \mid \langle\!\langle R \mid m : \varrho\rangle\!\rangle \mid \lambda\, t.R \mid R\varrho$$

- **Kinds**

$$\kappa ::= T \mid T^n \rightarrow [m_1, \ldots, m_k], \quad \text{pre} \quad (n \geq 0, k \geq 1).$$

The expression $\langle\!\langle R|m : \varrho\rangle\!\rangle$ extends the row R by the method named m, which is ϱ type and expression $\langle\!\langle\rangle\!\rangle$ indicates empty row, therefore the row, which does not contain any method.

The expression $R\varrho$ indicates the application of the row R to the ϱ type.

Rows and types are distinguished according to *kinds*. We can say that $[m_1, \ldots, m_k]$ components of the kind T are rows, that do not contain $m_1, \ldots, m_k$ methods. We need this information to statically ensure that methods would not multiple defined.

Row function $\lambda t.R$ will add object type to the row, which contains the names of the methods and types assigned to them. By object type adding it is allowed to the type of method depend on the type of host object. Row function $\lambda t.R$ has kinds in the form $T^n \rightarrow [m_1, \ldots, m_k]$. This kind form ensures, that if the function is applied to the type and the result (the resulting row) will not contain $m_1, \ldots, m_k$ methods.

System context consists of the type, row and kind context :

$$\Gamma ::= \Gamma, x : \varrho \mid \Gamma, t : T \mid \Gamma, R : \kappa$$

- $\Gamma \vdash x : \varrho$ - *type context*, each term x has ϱ type, at which value of the ϱ type is obtained by calculation of the term x;
- $\Gamma \vdash t : T$ - *kind context*, each type variable t is of the kind T;
- $\Gamma, R : \kappa$ - *row context*, each row R has assigned kind κ.

A. Typing rules

The most important part of type system are typing rules. In the text below we will mention only these rules which are relating to the object-oriented part :

1) Typing rule for empty object (pro $\langle\rangle$) :

$$\frac{}{\Gamma \vdash \langle\rangle : \textbf{pro } t.\langle\!\langle\rangle\!\rangle}.$$

Empty object does not include any methods, therefore can not respond to the sent messages. The type of empty object can be yield by **pro** $t.\langle\!\langle\rangle\!\rangle$ type expression. However the object does not include any methods, it can be extended by methods gained from any other object.

2) Typing rule for sending a message and method invocation (pro $\Leftarrow$) :

$$\frac{\Gamma \vdash e : \textbf{pro } t.\langle\!\langle R \mid m : \varrho\rangle\!\rangle}{\Gamma \vdash e \Leftarrow m : [\textbf{pro } t.\langle\!\langle R \mid m : \varrho\rangle\!\rangle/t]\varrho}.$$

If m method of the object e is invocated, it is claimed, that the **pro**type of object e would include method with the name m. The result of the method lookig up will be ϱ type and the **pro** $t.\langle\!\langle R \mid m : \varrho\rangle\!\rangle$ **pro**type of the host object will be substituted for type variable t. This substitution reflects recursive character of the **pro**types.

3) Typing rule for method adding and object extension (ext):

$$\Gamma \vdash \quad e_1 : \textbf{pro } t.\langle\!\langle \overline{m : \varrho}\rangle\!\rangle$$
$$\Gamma, t : T \vdash \quad R : [m_1, \ldots, m_k, n]$$

$$\frac{\Gamma, r : T \rightarrow [\overline{m}, n] \vdash e_2 : [\textbf{pro } t.\langle\!\langle rt \mid \overline{m : \varrho}, n : \varrho\rangle\!\rangle/t](t \rightarrow \varrho)}{\Gamma \vdash \langle e_1 \longleftrightarrow n = e_2\rangle : \textbf{pro } t.\langle\!\langle \overline{m : \varrho}, n : \varrho\rangle\!\rangle}.$$

Object e_1 can be extended by method n with a body e_2, if it holds :

- *first promise of the rule* - **pro**type of the object e_1 contains methods $m_1 : \varrho_1, \ldots, m_k : \varrho_k$. For

$m_1 : \varrho_1, \ldots, m_k : \varrho_k$ form simplification, in the text below we will write $\overline{m : \varrho}$;

- *second promise of the rule* - object e_1 does not include n method, which we are adding;
- *third promise of the rule* - e_2 body of the n method, has to be of a polymorphic function type. It ensures, that the body of the method will be performed by any extension of the host object, which includes n and $m_1, \ldots, m_k$ methods at least. The occurence of the row variable r with $T \to [\overline{m}, n]$ kind reflects to the claimed polymorphism in e_2 body type of the method n, which is applied to the type variable t. Type variable t represents polymorphic function type.

4) Typing rule for method redefining and object modification (mod) :

$$\Gamma \vdash e_1 : \textbf{pro } t.\langle\!\langle \overline{m : \varrho} \rangle\!\rangle$$

$$\frac{\Gamma, r : T \to [\overline{m}] \vdash e_2 : [\textbf{pro } t.\langle\!\langle rt \mid \overline{m : \varrho} \rangle\!\rangle / t](t \to \varrho)}{\Gamma \vdash \langle e_1 \leftarrow m_i = e_2 \rangle : \textbf{pro } t.\langle\!\langle \overline{m : \varrho} \rangle\!\rangle}.$$

IV. SUBTYPING

Subtyping [7], [8] is very simple, but many times used form of the polymorphism and it can be defined as substitution of **pro**type A object by **pro**type B object. Object of the **pro**type A can be used in any context as object of the **pro**type B if and only if object of the **pro**type A includes all methods of the **pro**type B object.

Because *method redefining* a *method adding* are allowed operations on objects, we distinguish two types of subtyping :

- **depth subtyping** ($\preceq$ depth) - is not able to use it, if *method redefining* operation is allowed;
- **width subtyping** (width $\preceq$) - is not able to use it, if *method adding* operation is allowed.

If it can be able to use subtyping in object calculus, we have to modify its syntax :

$\langle \rangle$	empty object;
$e \Leftarrow m$	sending message m to the object e;
$\langle e_1 \longleftarrow\!\!\lhd m = e_2 \rangle$	object e_1 extension by m with body e_2;
$\langle e_1 \leftarrow m = e_2 \rangle$	method m body e_1 redefine by body e_2;
$\langle e \longleftrightarrow m \rangle$	looking body of the method m in object e;
err	error object .

Modified production rule for object calculus we can defined as follows:

$$e ::= \quad x \mid c \mid \lambda x.e \mid e_1 e_2 \mid \langle \rangle \mid e \Leftarrow m \mid$$
$$\langle e_1 \longleftarrow\!\!\lhd m = e_2 \rangle \mid \langle e_1 \leftarrow m = e_2 \rangle \mid e \longleftrightarrow m \mid \text{err}.$$

Thus modified object calculus does not lose any characters of orginal object λ- calculus.

Grammar of modified object calculus is defined as follows:

- **Types**

$$\varrho ::= \underbrace{\varsigma}_{\text{type constants}} \mid \underbrace{t}_{\text{type variable}} \mid \underbrace{\varrho_1 \to \varrho_2}_{\text{function type}} \mid \underbrace{\textbf{pro } t.R}_{\textbf{pro}\text{totype}}$$

- **Labels**

$$\Delta ::= \{\} \mid \Delta \cup \{m\}$$

- **Labeled types**

$$\alpha ::= \varrho_\Delta$$

- **Rows**

$$R ::= \varrho \mid \langle\!\langle \rangle\!\rangle \mid \langle\!\langle R \mid m : \varrho \rangle\!\rangle \mid \lambda t.R \mid R\varrho$$

- **Kinds**

$$\kappa ::= T \mid T^n \to [m_1, \ldots, m_k], \quad \text{pre} \quad (n \geq 0, k \geq 1).$$

From thus defined grammar it results, that if ϱ is type and Δ is label, then ϱ_Δ is *labeled type*.

To prevent multiple method definition in modified object calculus, similarly like in orginal object calculus, are rows and types distinguished by kinds. Kind $[\overline{m_i}]$ includes rows, which do not include $\overline{m_i}$ methods.

Name set of row R, which is denoted as $\mathcal{S}(R)$, is defined as follows :

$\mathcal{S}(\langle\!\langle \rangle\!\rangle) = \{\}$	for empty row;
$\mathcal{S}(\langle\!\langle R \mid m : \varrho \rangle\!\rangle) = \mathcal{S}(R) \cup \Delta$	for extension of the row R by method with name m;
$\mathcal{S}(\lambda t.R) = \mathcal{S}(R)$	for function $\lambda t.R$, which add the type of the host object to the row;
$\mathcal{S}(R\varrho) = \mathcal{S}(R)$	for application the row R to the type ϱ.

System context contains terms, type and row variable :

$$\Gamma ::= \Gamma, x : \varrho \mid \Gamma, t : T \mid \Gamma, R : \kappa \mid \Gamma, \varrho_1 \preceq \varrho_2$$

- $\Gamma \vdash x : \varrho$ - *type context*, each term x has ϱ type, at which value of the ϱ type is obtained by calculation of the term x;
- $\Gamma \vdash t : T$ - *kind context*, each type variable t is of the kind T;
- $\Gamma, R : \kappa$ - *row context*, each row R has assigned kind κ;
- $\Gamma, \varrho_1 \preceq \varrho_2$ - subtyping relation.

A. Subtyping relation

Restricted form of the *subtyping* in modified object calculus allows using (width $\preceq$), or (depth $\preceq$) subtyping together with method adding, respectively method redefining.

This restriction has to be valid :

- object of the some **pro**type, for example A can be subtyped by the object of the any other **pro**type, for example B if and only if we ensure, that methods of the **pro**type A object, that are not of the **pro**type B object, are not of the **pro**type B object methods either requested, therefore methods must either not refer to them.

If this restriction was not valid and methods of the **pro**type B object would refer to the **pro**type A object methods, it could cause error during program running.

Subtyping relation allows to hide (forget) the method in object without a change of the object form. Information about which method is used is collected by implementation of the *labeled types*.

In the text below we will introduce only rules for (width $\preceq$) subtyping, (sub $\preceq$) subtyping and ($\preceq$ depth) subtyping. List of all rules is presented in [6].

1) (width $\preceq$) subtyping rule

$$\frac{\Gamma \vdash \textbf{pro } t.\langle\!\langle R \mid \overline{n : \alpha} \rangle\!\rangle : T \qquad \overline{n} \notin \mathcal{S}(R)}{\Gamma \vdash \textbf{pro } t.\langle\!\langle R \mid \overline{n : \alpha} \rangle\!\rangle}.$$

(Width $\preceq$) subtyping rule denotes, that type is subtype of other type, if hidden (forgotten) methods in the other type are not in association with labeled sets of remaining

methods. Condition validity $\overline{n} \notin \mathcal{S}(R)$ ensures, that remaining methods are not using $\overline{n}$ methods.

2) (sub $\preceq$) subtyping rule

$$\frac{\Gamma \vdash e : \varrho_1 \quad \Gamma \vdash \varrho_1 \preceq \varrho_2}{\Gamma \vdash e : \varrho_2}$$

(Sub $\preceq$) subtyping rule allows to assign more types to some object and is able to use it for constants subtyping. This subtyping rule presents, that if object e is ϱ_1 type and if it holds $\Gamma \vdash \varrho_1 \preceq \varrho_2$, then for object e we also can derive ϱ_2 type. It follows, that object of the ϱ_1 type can be used in any context, in which ϱ_2 object type is desired.

3) ($\preceq$ depth) subtyping rule

$$\frac{\Gamma \vdash e : \mathbf{pro}\ t.\langle\!\langle R \mid n : \varrho_\Delta \rangle\!\rangle : T \quad \Delta \subseteq \Delta'}{\Gamma \vdash e : \mathbf{pro}\ t.\langle\!\langle R \mid n : \varrho_{\Delta'} \rangle\!\rangle}.$$

Example 4.1 (Subtyping between objects): Let's have two objects :

$$point_1 \stackrel{def}{=} \quad \langle\ x = \langle \lambda self.0,$$
$$move_x = \lambda self.\lambda feed_x.\langle self \leftarrow x = \lambda s.(self \Leftarrow x) + move_x \rangle$$
$$\rangle,$$
$$point_2 \stackrel{def}{=} \quad \langle\ x = \lambda self.1, y = \lambda self.0,$$
$$move_x = \lambda self.\lambda feed_x.\langle self \leftarrow x = \lambda s.(self \Leftarrow x) + feed_x, \rangle$$
$$move_y = \lambda self.\lambda feed_y.\langle self \leftarrow y = \lambda s.(self \Leftarrow y) + feed_y \rangle$$
$$\rangle,$$

where $point_1$ is of Point$_1$ type and $point_2$ is of Point$_2$ type :

$$\text{Point}_1 \stackrel{def}{=} \quad \mathbf{pro}\ t.\langle\!\langle \text{x:int}, move_x : (int \rightarrow t)_x \rangle\!\rangle,$$
$$\text{Point}_2 \stackrel{def}{=} \quad \mathbf{pro}\ t.\langle\!\langle \text{x:int,y:int}, move_x : (int \rightarrow t)_x, move_y : (int \rightarrow t)_y, \rangle\!\rangle.$$

For thus defined objects we can say, that Point$_2 \preceq$ Point$_1$, because y method can be hidden.

If we modify the objects :

$$point_1' \stackrel{def}{=} \quad \langle point_1 \leftarrow move_x = \lambda self.\lambda \text{feed}_x.self \rangle,$$
$$point_2' \stackrel{def}{=} \quad \langle\!\langle point_2 \leftarrow x = \lambda self.(self \Leftarrow move_x\ 1) \rangle\!\rangle \leftarrow move_x =$$
$$\lambda self.\lambda \text{feed}_x, self \rangle,$$

where :

$$\text{Point}_1' \stackrel{def}{=} \quad \mathbf{pro}\ t.\langle\!\langle \text{x:int}, move_x : (int \rightarrow t)_x \rangle\!\rangle,$$
$$\text{Point}_2' \stackrel{def}{=} \quad \mathbf{pro}\ t.\langle\!\langle \text{x:int}_{\{move_x, y\}}, \text{y:int}, move_x : (int \rightarrow t), move_y : (int \rightarrow t)_{\{y\}}, \rangle\!\rangle.$$

In this case we can say, that Point$_2' \npreceq$ Point$_1'$, because we cannot hide (forget) y method. It is used by x method. For this reason we cannot assign Point$_1'$ type to the $point_2'$ object. If we did it, it would cause error during program running. This way we will avoid error message.

$\square$

V. Conclusion

Object calculus, especially type object calculus has some interesting futures for behaviour description of the software objects and represent appropriate mechanism for modelling of software objects. Behaviour description of the software object is the main reason of the type object calcul definition.

Type of the software object is defined by type expression **pro** $t.R$, that is named **pro**type and describes all software objects which contains methods m_i, that the result of message sending to the object ($e \Leftarrow m_i$) is value of ϱ_i type. Messages sending is the main operation between software objects. Software objects communicate by sending of the messages what is in type object calculus described by "sealing" a **pro**type into a real object type, therefore internal methods can be overrided by other methods of their host object.

If we can assign a **pro**type to an software object, then we can modify the behaviour of the software object by object extension or object modification. Object extension and object modification are abilities of software object behaviour modification by addinng new methods and inheriting existing ones, respectively by redefining existing methods.

Behaviour of the software object can be modified by subtyping, too. Object calculus supports two types of restricted subtyping because we cannot use subtyping when method adding or method redefinig operations are allowed. In this regard has to be valid restriction that object of the some **pro**type, for example A can be subtyped by the object of the any other **pro**type, for example B if and only if we ensure, that methods of the **pro**type A object, that are not of the **pro**type B object, are not of the **pro**type B object methods either requested. If this restriction was not valid it could cause error during program running.

In our next research we would like to focus on modeling software object in category theory and defining suitable categorical structures for coalgebraic behaviour description of object-oriented and complex program systems. We would like to apply achieved theoretical results to real non trivial object-oriented or complex program systems from the area of computer networks, as for example intrusion detection system [9] or database systems.

References

[1] K. S. Fisher, F. Honsell, and J. C. Mitchell, "A Lambda Calculus of Objects and Method Specialization," *Nordic Journal of Computing.*, pp. 1(1):3–37, 1994.

[2] L. Liquori and G. Castagna, "A Typed Lampda Calculus of Objects."

[3] A. Church, "A set of postulates for the foundation of logic," *Annals of Mathematics*, pp. 346–366, 1932.

[4] R. Rojas, "A Tutorial Introduction to the Lambda Calculus," free University of Berlin, WS-97/98.

[5] K. S. Fisher, "Type systems for object-oriented programming languages," Ph.D. dissertation, Stanford university, 1996.

[6] V. Bono, "Type system for the object-oriented paradigm," Ph.D. dissertation, Universita di Torino, 1996.

[7] M. Jenčik, "Formálny popis správania objektovo-orientovaných programov," písomná práca k dizertačnej skúške, KPI FEI TUKE Košice, 2010.

[8] V. Bono and L. Liquori, "A Subtyping for the Fisher-Honsell-Mitchell Lambda Calculus of Objects." in *Proceedings of of CSL, volume 933 of Lecture Notes in Computer Science*. Springer-Verlag, 1995, pp. 16–30.

[9] D. Mihályi and V. Novitzká, "A Coalgebra as an Intrusion Detection System," *Acta Polytechnica Hungarica, Journal of Applied Sciences, O´buda University*, vol. 7, no. 2, pp. 71–79, 2010.

On the join products with crossing number one

Jana PETRILLOVÁ

Dept. of Mathematics and Theoretical Informatics, FEI TU of Košice, Slovak Republic

jana.petrillova@tuke.sk

Abstract—**Kulli at al. started to characterize line graphs with crossing number one. In this paper, the similar problem was solved for the join product of two graphs. The necessary and sufficient conditions are given for all pairs of graphs G_1 and G_2, both with at least one edge, for which the crossing number of their join product $G_1 + G_2$ is one.**

Keywords—**join product, crossing number, drawing, graph**

I. INTRODUCTION

Let G be a simple graph with vertex set V and edge set E. A *drawing* of a graph is a mapping of a graph into a surface. For simplicity, we assume that in a drawing (a) no edge passes through any vertex other than its end-points, (b) no two edges touch each other (i.e., if two edges have a common interior point, then at this point they properly cross each other), and (c) no three edges cross at the same point. It is easy to see that a drawing with minimum number of crossings, an *optimal* drawing, is always a good drawing, meaning that no edge crosses itself, no two edges cross more than once, and no two edges incident with the same vertex cross. The *crossing number* $cr(G)$ of a graph G is the number of edge crossings in an optimal drawing of G in the plane.

Let G_1 and G_2 be simple graphs with vertex sets $V(G_1)$ and $V(G_2)$, and edge sets $E(G_1)$ and $E(G_2)$. The join product of two graphs G_1 and G_2, denoted by $G_1 + G_2$, is obtained from vertex-disjoint copies of G_1 and G_2 by adding all edges between $V(G_1)$ and $V(G_2)$. For $|V(G_1)| = m$ and $|V(G_2)| = n$, the edge set of $G_1 + G_2$ is the union of disjoint edge sets of the graphs G_1, G_2, and the complete bipartite graph $K_{m,n}$. Let us denote by C_n and P_n the cycle and the paths with n vertices, respectively, and let S_n be the star $K_{1,n}$.

The crossing numbers has been studied to improve the readability of hierarchical structures and by VLSI communities and computer scientists. As a crossing of two edges of the communication graph requires unit area in VLSI-layout, the crossing number together with the number of vertices of the graph immediately provide a lower bound for the area of the VLSI-layout of the communication graph.

As computing the exact value of crossing number of a given graph is in general an elusive problem, the crossing numbers of few families of graphs are known. Very important result, which is often used to find the crossing number of join product of two graphs, is known Kleitman's result on crossing numbers of complete bipartite graphs. More precisely, he proved that for $\min\{m,n\} \leq 6$

$$cr(K_{m,n}) = Z(m,n) = \left\lfloor \frac{m}{2} \right\rfloor \left\lfloor \frac{m-1}{2} \right\rfloor \left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor.$$

Another important results, which are used in this paper, are the crossing numbers of join product of two paths, join product

of two cycles, and join product of path and cycle, see [3]. The exact values of crossing numbers of $G + P_n$ and $G + C_n$ for all graphs G of order at most four are given in [5], and the crossing numbers of join of the special graph on six vertices with path and cycle were collected in [4].

Kulli at al. in [6] started to study line graphs with crossing number one. In [7], Kulli and Muddebihal gave the characterization for all pairs of graphs for which the crossing number of their join product is zero. In this paper, we give the necessary and sufficient conditions for all pairs of graphs G_1 and G_2, both with at least one edge, for which the crossing number of their join product $G_1 + G_2$ is one.

II. PRELIMINARY RESULTS

Let us consider five graphs in Fig. 1. In the proof of the main result of the paper, we need to know the crossing numbers of the graphs $J_1 + P_2$, $J_2 + P_2$, $J_3 + P_2$, $J_4 + P_2$ and $J_5 + P_2$. The crossing number of the graphs $J_4 \times P_2$ was established in [3]. The other results are collected in the next lemmas. Let x and y denote the vertices of the path P_2.
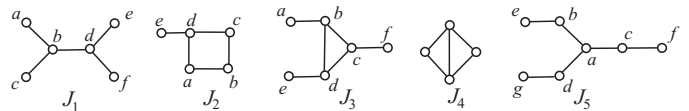
Fig. 1. The graphs J_1, J_2, J_3, J_4 and J_5.

Lemma 1 $cr(J_1 + P_2) = 2$

Proof. Fig. 2(a) shows that $cr(J_1 + P_2) \leq 2$. The graph J_1 consists of the graph S_3 with the vertices a, b, c, d and of the path edf, see Fig. 1. As the graph $J_1 + P_2$ contains $S_3 + P_2$ as a subgraph and $cr(S_3 + P_2) = 1$ (see [3]), $1 \leq cr(J_1 + P_2)$. If there is a drawing D of the graph $J_1 + P_2$ with one crossing, then none of the edges incident with the vertices e and f is crossed in D. Let J_1^{ef} be the subgraph of the graph $J_1 + P_2$ induced by the edges incident with the vertices e and f. Clearly, J_1^{ef} is isomorphic with the graph $K_{2,3}$ (see Fig. 2(b)). The vertex b is adjacent with the vertices d, x, and y in the subgraph $S_3 + P_2$. Since the graph $K_{2,3}$ is not outerplanar graph and the edges of J_1^{ef} do not cross each other in D, the edges joining the vertex b with the vertices d, x, and y cross the edges of J_1^{ef} at least once. This contradiction with the restriction of no crossing on the edges incident with the vertices e and f forces that $cr(J_1 + P_2) \geq 2$. Hence, $cr(J_1 + P_2) = 2$, and the proof is done. $\square$

Lemma 2 $cr(J_2 + P_2) = 2$

(a)   (b)
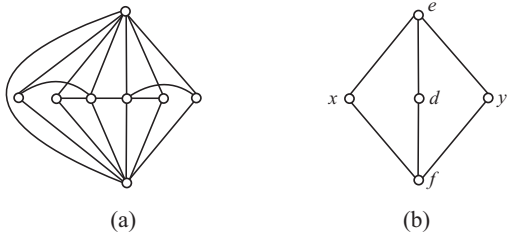
Fig. 2.   The graph $J_1 + P_2$ with two crossings and its subgraph J_1^{ef}.

Proof. Fig. 3(a) shows that $cr(J_2 + P_2) \leq 2$. The graph J_2 consists of the 4-cycle C_4 with the vertices a, b, c, d and of the edge de, see Fig. 1. The graph $J_2 + P_2$ contains $C_4 + P_2$ as a subgraph and therefore, $1 = cr(C_4 + P_2) \leq cr(J_2 + P_2)$ (see [3]). If there is a drawing D of the graph $J_2 + P_2$ with one crossing, then none of the edges joining the vertices x and y with the 4-cycle is crossed in D. Otherwise, by deleting a crossed edge joining x or y with the 4-cycle the subdrawing without crossings is obtained. This is impossible, because the resulting subgraph contains a subdivision of K_5 and $cr(K_5) = 1$. Thus, the edges joining x and y with the vertices of the 4-cycle must be drawn as in Fig. 3(b). In Fig. 3(c) there is shown the unique subdrawing of $C_4 + P_2$ obtained from the drawing in Fig. 3(b) by adding the edges of the 4-cycle without crossings. Let J_2^e be the subgraph of the graph $J_2 + P_2$ induced by the edges incident with the vertex e. It is easy to verify that, in D, the edges of J_2^e cross the edges of $C_4 + P_2$ shown in Fig. 3(c). This contradicts the assumption of at most one crossing in D. Hence, $cr(J_2 + P_2) = 2$.   □

(a)   (b)   (c)
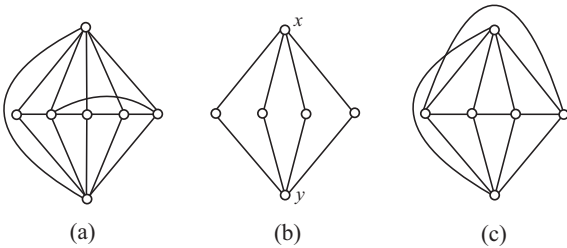
Fig. 3.   The graph $J_2 + P_2$ with two crossings and the necessary subdrawings of $C_4 + P_2$.

Lemma 3 $cr(J_3 + P_2) \geq 2$

Proof. The graph J_3 consists of the graph C_3 with the vertices b, c, d and of the edges ab, cf, and de, see Fig. 1. The graph $J_3 + P_2$ contains $C_3 + P_2$ as a subgraph and therefore, $1 = cr(C_3 + P_2) \leq cr(J_3 + P_2)$ (see [3]). If there is a drawing D of the graph $J_3 + P_2$ with one crossing, then none of the edges incident with the vertices a, e and f is crossed in D. Let J_3^{aef} be the subgraph of the graph $J_3 + P_2$ consisting of the edges joining the vertices a, e and f with the vertices x and y. Clearly, J_3^{aef} is isomorphic with the graph $K_{2,3}$ (see Fig. 4(a)). Since the graph $K_{2,3}$ is not outerplanar graph and the edges of J_3^{aef} do not cross each other in D, the edge ab or one of the paths bde and bcf cross the edges of J_3^{aef} at least once. This confirms that there is no drawing of the graph $J_3 + P_2$ with only one crossing. So, $cr(J_3 + P_2) \geq 2$.
    □

Lemma 4 $cr(J_5 + P_2) = 2$

Proof. The graph J_5 consists of the graph S_3 with the vertices a, b, c, d and of the edges dg, cf, and be, see Fig. 1. As the graph $J_5 + P_2$ contains $S_3 + P_2$ as a subgraph, $1 = cr(S_3 + P_2) \leq cr(J_5 + P_2)$ (see [3]). Consider a drawing D of the graph $J_5 + P_2$ with one crossing. Then none of the edges incident with the vertices e, f and g is crossed in D. Let J_5^{efg} be the subgraph of the graph $J_5 + P_2$ consisting of the edges joining the vertices e, f and g with the vertices x and y. Clearly, J_5^{efg} is isomorphic with the graph $K_{2,3}$ (see Fig. 4(b)). Since the graph $K_{2,3}$ is not outerplanar graph and the edges of J_5^{efg} do not cross each other in D, at least one of the paths abe, adg, and acf cross the edges of J_5^{efg} in D. Hence, $cr(J_5 + P_2) \geq 2$. The reverse inequality $cr(J_5 + P_2) \leq 2$ follows from the drawing in Fig. 4(c). This completes the proof.   □

(a)   (b)   (c)
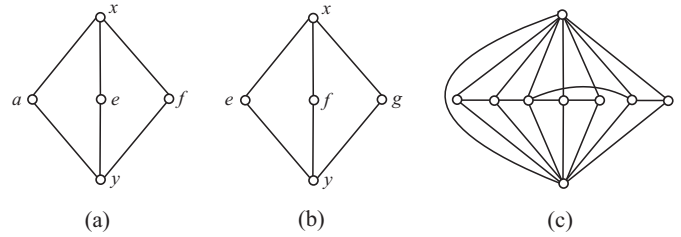
Fig. 4.   The subgraphs J_3^{aef}, J_5^{efg} and the graph $J_5 + P_2$ with two crossings.

III. THE MAIN RESULT

The aim of this section is to characterize all pairs of graphs on at least one edge for which their join product have crossing number one. We will consider the next graphs: the star S_3, the graphs S_{3+} and S_{3++} obtained by subdividing one and two edges of S_3, respectively, the graph H_1 shown in Fig. 5, the graph H_{1+} obtained from H_1 by subdividing the non-cycle edge, the graph H_2 shown in Fig. 5, and the graphs H_{2+} and H_{2++} obtained from H_2 by subdividing one or both non-cycle edges, respectively. Let G_p be the graph whose components are the paths P_n for $n \geq 1$. We will consider also the graphs $S_3 \cup G_p$, $S_{3+} \cup G_p$, $S_{3++} \cup G_p$, $H_1 \cup G_p$, $H_{1+} \cup G_p$, $H_2 \cup G_p$, $H_{2+} \cup G_p$, $H_{2++} \cup G_p$, and $C_3 \cup G_p$.
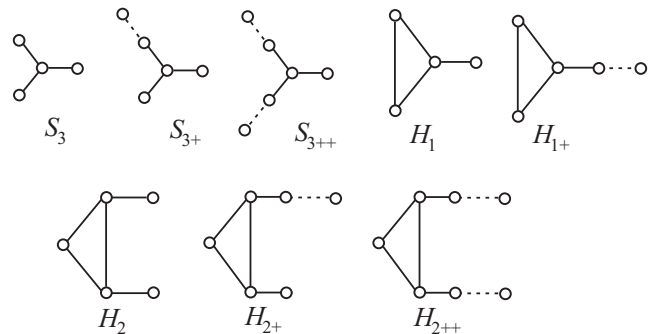
Fig. 5.   The connected graphs used for characterization of join products with crossing number one.

Theorem 1 *Let G_1 and G_2 be two graphs, both with at least one edge. Then $cr(G_1 + G_2) = 1$ if and only if one of the following conditions holds:*

    a) $G_1 = P_3$ *and G_2 is one of $P_1 \cup P_2$ or P_3,*
    b) $G_1 = P_1 \cup P_2$ *and G_2 one of $P_1 \cup P_2$ or C_3,*

c) $G_1 = P_2$ and G_2 is one of S_3, S_{3+}, S_{3++}, H_1, H_{1+}, H_2, H_{2+}, H_{2++}, $S_3 \cup G_p$, $S_{3+} \cup G_p$, $S_{3++} \cup G_p$, $H_1 \cup G_p$, $H_{1+} \cup G_p$, $H_2 \cup G_p$, $H_{2+} \cup G_p$, $H_{2++} \cup G_p$, $C_3 \cup G_p$ or C_n for $n \geq 3$.

Proof. It was proved in [3] that $cr(P_m + P_n) = Z(m,n)$ for $m \geq 1$, $n \geq 1$, $\min\{m,n\} \leq 6$ and that $cr(P_m + C_n) = Z(m,n) + 1$ for $m \geq 2$, $n \geq 3$, $\min\{m,n\} \leq 6$. This implies that $cr(P_3 + P_3) = cr(P_2 + C_n) = 1$ for $n \geq 3$. It was also proved in [3] that $cr(P_2 + S_3) = cr((P_1 \cup P_2) + P_3) = cr((P_1 \cup P_2) + C_3) = 1$ for $n \geq 3$. The graph $(P_1 \cup P_2) + (P_1 \cup P_2)$ contains $K_{3,3}$ as a subgraph. As $cr(K_{3,3}) = 1$, the crossing number of $(P_1 \cup P_2) + (P_1 \cup P_2)$ is at least one. The reverse inequality follows from the drawing in Fig. 6(a). All graphs S_{3+}, S_{3++}, H_1, H_{1+}, H_2, H_{2+}, H_{2++}, $S_3 \cup G_p$, $S_{3+} \cup G_p$, $S_{3++} \cup G_p$, $H_1 \cup G_p$, $H_{1+} \cup G_p$, $H_2 \cup G_p$, $H_{2+} \cup G_p$, and $H_{2++} \cup G_p$ contain S_3 as a subgraph and therefore, the join product of these graphs with the path P_2 contain $P_2 + S_3$. This implies that if $G_1 = P_2$ and G_2 is one of S_{3+}, S_{3++}, H_1, H_{1+}, H_2, H_{2+}, H_{2++}, $S_3 \cup G_p$, $S_{3+} \cup G_p$, $S_{3++} \cup G_p$, $H_1 \cup G_p$, $H_{1+} \cup G_p$, $H_2 \cup G_p$, $H_{2+} \cup G_p$ or $H_{2++} \cup G_p$, then $cr(G_1 + G_2) \geq 1$. As the graphs S_{3+}, S_{3++}, and $S_{3+} \cup G_p$ are subgraphs of the graph $S_{3++} \cup G_p$ and also the graphs H_1, H_{1+}, H_2, H_{2+}, H_{2++}, $H_1 \cup G_p$, $H_{1+} \cup G_p$, $H_2 \cup G_p$, $H_{2+} \cup G_p$, and $S_{3+} \cup G_p$ are all subgraphs of the graph $H_{2++} \cup G_p$, the reverse inequality $cr(G_1 + G_2) \leq 1$ follow from the drawings in Fig. 6(b) and Fig. 7. The graph $C_3 \cup G_p$ is also a subgraph of the graph $H_{2++} \cup G_p$ and, as $cr(C_3 + P_2) = 1$, we have that $cr((C_3 \cup G_p) + P_2) = 1$.

(a)   (b)
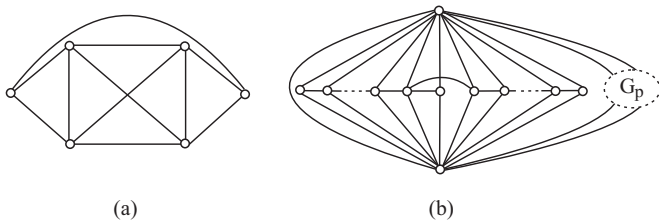
Fig. 6.   The graphs $(P_1 \cup P_2) + (P_1 \cup P_2)$ and $(S_{3++} \cup G_p) + P_2$ with one crossing.
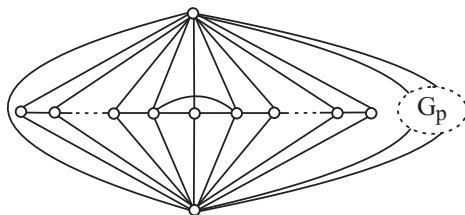
Fig. 7.   The graph $(H_{2++} \cup G_p) + P_2$ with one crossing.

It remains to prove that there are no other graphs G_1 and G_2 for which the crossing number of $G_1 + G_2$ is one, when both G_1 and G_2 contain at least one edge. Assume that $cr(G_1 + G_2) = 1$. Clearly, at most one of the graphs G_1 and G_2 contain more than three vertices. Otherwise the graph $G_1 + G_2$ contains $K_{4,4}$ as a subgraph and $cr(K_{4,4}) = 4$. Let G_1 contains at most three vertices. Then G_1 is one of the following graphs: P_2, $P_1 \cup P_2$, P_3 or C_3.

Assume first that $G_1 = P_3$. Then G_2 contains also at most three vertices, otherwise the graph $G_1 + G_2$ contains $K_{4,3}$ as a subgraph and $cr(K_{4,3}) = 2$. It is easy to see that graph $P_3 + P_2$ is planar. If $G_2 = C_3$, then $cr(G_1 + G_2) = cr(P_3 + C_3) =$

$Z(3,3) + 1 = 2$, see [3]. Thus, G_2 is one of P_3 or $P_1 \cup P_2$ and the crossing number of $G_1 + G_2$ is one.

If $G_1 = C_3$, then G_2 contains at most three vertices, see above. As $cr(G_1 + P_3) = cr(C_3 + P_3) = 2$ and $cr(G_1 + C_3) = cr(C_3 + C_3) = 3$ (see [3]), G_2 is one of $P_1 \cup P_2$ or P_2.

If $G_1 = P_2$, then G_2 contains vertices of degree at most three. Otherwise the graph $G_1 + G_2$ contains $K_{4,3}$ as a subgraph. Assume first that G_2 contains only vertices of degree at most two. As $cr(P_2 + P_n) = Z(2,n) = 0$, G_2 cannot be a path. It is easy to see, that also the graph $P_2 + G_p$ is planar. This implies that $G_2 = C_n$, because $cr(P_2 + C_n) = Z(2,n) + 1 = 1$ for $n \geq 3$ (see [3]). Moreover, G_2 can be the graph $C_3 \cup G_p$. Let the maximum degree of the graph G_2 be three. If G_2 contains only one vertex of degree three, which is not contained in a cycle of G_2, then G_2 is one of S_3, S_{3+}, S_{3++}, $S_3 \cup G_p$, $S_{3+} \cup G_p$, and $S_{3++} \cup G_p$. The graph G_2 cannot be a subdivision of S_3 on all three leaves, otherwise G_2 contains J_5 as a subgraph and, by Lemma 4, $cr(P_2 + J_5) = 2$. If the unique vertex of degree three is contained in a cycle, then it is contained in G_2 in a cycle of length three, otherwise G_2 contains a subdivision of J_2 as a subgraph. In this case, by Lemma 2, $cr(P_2 \times G_2) > 1$. Thus, G_2 is one of H_1, $H_1 \cup G_p$, H_{1+} or $H_{1+} \cup G_p$. Assume now that G_2 contains more than one vertex of degree three. Then all of them are contained in a cycle of G_2, because otherwise a graph homeomorphic with the graph J_1 is a subgraph of G_2 and, by Lemma 1, $cr(P_2 + J_1) = 2$. By Lemma 3, $cr(P_2 + J_3) \geq 2$ and $cr(P_2 + J_4) = 2$ (see [3]). Thus, G_2 does not contain more than two vertices of degree three, otherwise G_2 contains a subgraph homeomorphic to the graph J_3 and $cr(P_2 + G_2) > 1$. Both vertices of degree three are contained in G_2 in a cycle of length three, otherwise G_2 contains a subdivision of J_2 as a subgraph. Hence, if there are more than two edge-disjoint paths between the vertices of degree three in the graph G_2, then G_2 contains J_4 as a subgraph and $cr(P_2 + G_2) \geq 2$. If there are only two edge-disjoint paths between the vertices of degree three in G_2, then G_2 is one of H_2, $H_2 \cup G_p$, H_{2+}, $H_{2+} \cup G_p$, H_{2++} or $H_{2++} \cup G_p$. If G_2 is disconnected graph and the vertices of degree three are not in the same component, then G_2 contains at most one vertex of degree three. Otherwise G_2 contains two edge-disjoint subgraphs $K_{1,3,2}$ and Asano proved that $cr(K_{1,3,n}) = Z(4,n) + \lfloor \frac{n}{2} \rfloor$. This implies that $cr(P_2 + G_2) \geq cr(K_{1,3,2}) + cr(K_{1,3,2}) \geq 2$.

If $G_1 = P_1 \cup P_2$, then G_2 contains at most three vertices. Moreover, G_2 contains the vertices of degree at most two. This implies that G_2 contains only paths and cycles. It is easy to see that the graphs $G_1 + P_2$ is planar. Klešč in [3] proved that $cr((P_1 \cup P_2) + P_3) = cr((P_1 \cup P_2) + C_3) = 1$. If G_2 is disconnected graph, then G_2 is also $P_1 \cup P_2$. This completes the proof.   □

IV. CONCLUSION

The investigation on the crossing numbers of graphs is very difficult problem. In this paper, we proved the necessary and sufficient conditions for all pairs of graphs G_1 and G_2 for which the crossing number of their join product $G_1 + G_2$ is one, but only if both graphs have at least one edge. It remains to prove the cases, when some of the graphs G_1 and G_2 is the path P_1 or none of G_1 and G_2 contains an edge.

REFERENCES

[1] K. Asano, The crossing number of $K_{1,3,n}$ and $K_{2,3,n}$, *J. Graph Theory* **10** (1986), 1–8.

[2] D. J. Kleitman, The crossing number of $K_{5,n}$, *J. Combin. Theory Ser. B* **9** (1971), 315-323.

[3] M. Klešč, The join of graphs and crossing numbers, *Electronic Notes in Discrete Mathematics* **28** (2007), 349–355.

[4] M. Klešč, The crossing numbers of join of the special graph on six vertices with path and cycle, *Discrete Math.* **310** (2010), 1475–1481.

[5] M. Klešč and Š. Schrötter, The crossing numbers of join products of paths with graphs of order four, *Discussiones Mathematicae - Graph Theory*, (to be published).

[6] V. R. Kulli, D. G. Akka, L. W. Beineke, On line graphs with crossing number one, *J. Graph Theory* **3** (1979), 87–90.

[7] V. R. Kulli and M. H. Muddebihal, Characterization of join graphs with crossing number zero, *Far East J. Appl. Math.* **5** (2001), 87-97.

Overview of Haskell Syntax Analyzer as a Tool to Create Haskell Domain-Specific Dialect

Ľubomír WASSERMANN, Emília PIETRIKOVÁ, Sergej CHODAREV

Dept. of Computers and Informatics, Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic
Lubomir.Wassermann@tuke.sk, Emilia.Pietrikova@tuke.sk, Sergej.Chodarev@tuke.sk

Abstract—**General purpose languages offer developers large set of constructs to develop software systems from various domains. It results in general usability of these languages but makes them more complex. If we want to reach higher development productivity in a specific domain, it is suitable to have domain-specific dialect of the programming language. This paper gives an overview about Haskell Syntax Analyzer tool. The goal of this tool is to analyze programs written in Haskell language. Based on the analysis of programs and their derivation trees, we can obtain the set of used constructs (grammar rules) in programs from particular domain. Then we can reduce the set of language constructs to those which are used and related with the domain. Deeper analysis of gathered data can offer various statistics focused on presence of language constructs or to recognize and apply language patterns in Haskell programs.**

Keywords—**domain-specific dialect, Haskell syntax analyzer, parser generator, program analyzer**

I. INTRODUCTION

During the development of a software system, two domains are meeting. Problem domain, which is represented by domain expert[1]. On the other hand, the solution domain, which is represented by a software system to be developed itself.

Description of the problem called requirements specification is expressed in the language of the problem domain. The language of the solution domain is a programming language chosen by the programmer. It represents a tool that can describe the solution of the problem from particular domain in a form understandable to computer. Since these languages are very different, finding a solution to the problem is a complex process. The actual creation of a software system can thus be defined as the concepts transformation of the problem domain into concepts of the solution domain.

In software engineering, various trends and approaches emerged how to improve the existing situation. In some cases, use of general-purpose languages (GPL) is not appropriate to solve specific problems. This widens the gap between the problem and the solution domain due to differences of domain languages that results in communication barrier between domain expert and system developer. A possible solution is to bring the problem, solution domains and their respective languages closer. Therefore, the importance and popularity gained a new approach to software development named language-oriented programming. This approach relies on creation of own small languages of high abstraction oriented to a specific domain. These languages are called domain-specific languages.

---

[1] *Domain expert is a person who has practical experience in the domain of the problem and the necessary knowledge in this domain needed to solve the problem.*

The idea of the proposed tool is to process Haskell programs and provide appropriate derivation trees according to pre-processed Haskell grammar [1]. Based on the analysis of the derivation trees, we can decide which language constructs are used in that particular domain. Thus, we can reduce the set of language constructs to create a domain-specific dialect of the language.

II. DOMAIN-SPECIFIC LANGUAGES

Domain-specific languages (DSL) are computer languages aimed at solving problems from particular domain. They are essentially small languages with limited expressiveness, offering programmer limited set of concepts, notations and abstractions from the target domain [2]. Thanks to the higher level of abstraction by using the domain concepts, the productivity of programmers and their communication with the domain experts is improved when compared to using GPL [3]. When design of DSL is simple and suitable for the domain, it is even possible that DSL can be used by the domain expert or end user (end-user programming [2]).

A. Pros & Cons of domain-specific languages

The main advantages of using DSLs are following [2], [4], [5]:

- *Capturing domain knowledge* - DSL uses concepts of the problem domain and defines their mutual relations. DSL represents the domain knowledge and allows their reuse in other projects and maintains them in a consistent state.
- *Participation of domain expert* - Thanks to the use of concepts from the language of domain, the domain expert can easily understand, validate, edit or even write DSL program itself. Improving communication with the domain expert and his closer involvement in the software development process helps prevent errors in the design and implementation of requirements.
- *Improving productivity* - Thanks to the use of higher level abstractions, DSLs improve productivity of programmers. DSL programs are shorter, more readable, easier to mantain and less prone to errors. Since language of the problem domain is closer to the language of the solution domain, it is easier to write a solution of the problem as opposed to the use of GPL.
- *Reusability* - If a DSL is well designed, its reusability presents an advantage. Due to increased costs associated with an implementation of the DSL, it is not suitable to create and use DSL for single solutions. However, coping

with more same or similar problems, use of appropriate DSL brings great benefits.

- *Changing Paradigms* - Using DSL to solve problems can help to overcome deficiencies of GPL and find simplier solution of the problem with an appropriate DSL, which is designed for this domain [6].

On the other hand, the use of DSL also brings disadvantages. However, most of these shortcomings can be addressed to the fact that the use of DSL is not yet established and most programmers do not know how to properly design and implement DSL or do not see the importance of DSL on a wider scale in software development [6]. The disadvantages of DSL are:

- *Learning curve* - Every new technology brings additional costs with it because it is necessary to get familiar with its principles and gain experience how to apply them correctly. Gained knowledge can be subsequently applied to other projects but the initial use of new technology is connected with higher costs [6].
- *Cost of building* - Despite the knowledge of DSL technology, additional costs for creating the DSL are needed. Good design, implementation and maintenance of DSL requires knowledge and experience in language construction. When designing a DSL, it is necessary to keep in mind that DSL should have only limited expressiveness [6].
- *Lack of tool support* - There is no existing widespread, fully functional, integrated environment to support the construction and use of DSL in the development of software systems [4]. Several research groups and development teams are working on the environment with such a support as described in [6]–[8].
- *Difficulty of migrating DSL programs* - If it is necessary to change DSL due to new requirements, it rises a problem of migration and compatibility issues of already created programs. Due to the absence of a tool support for the evolution of DSL, it is necessary to ensure the migration or to avoid it with support of older versions of the DSL [6].
- *Loss of domain specificity* - During its life cycle, DSL is a subject to a number of changes that leads to the evolution of DSL. With longer use of DSL, there is a risk of slipping to generality and specialized DSL can become language with expressiveness of GPL [2], [9].

III. HASKELL SYNTAX ANALYZER TOOL

During development of software systems, programmers mostly use GPL languages. But when they are dealing with problems from same domain, they possibly do not need the whole set of constructs and syntax elements offered by complex GPL. It is more suitable to create subset of language constructs that are used in particular domain and create lightweight version of language as its dialect. This can lead to optimization of the language and its compiler for particular domain.

As the diversity of programming paradigms is perspicuous, the importance of domain-specific notation in programming languages is increasingly accepted [10]. It becomes more important to be able to try out new language features and new notation. As an extension of the Haskell language that does not change its intrinsic nature, a dialect might be created

for use in a domain-specific language. Ideally, we should be able to process the existing language features in order to benefit from user experience before full scale production implementation and avoid expensive modifications to the new language implementation later.

The aim of creating *Haskell Syntax Analyzer tool* is to gather needed information from Haskell programs to get a proper knowledge about used constructs in analyzed programs. The result of program analysis is a derivation tree consisting of used rules of Haskell grammar. Architecture of Haskell Syntax Analyzer tool consists of two parts - *generating infrastructure* and *analyzing infrastructure*. The goal of generating infrastructure is to prepare tools which are used during analysis of Haskell programs by analyzing infrastructure. The analyzing infrastructure consists of lexer and parser of Haskell programs. Result of analysis are the derivation trees of the analyzed programs. The whole architecture of Haskell Syntax Analyzer tool can be seen in fig. 1.

A. Generating infrastructure

As already mentioned, the main goal of generating infrastructure is to prepare tools used in program analysis. This tools (lexer, parser) have to conform to Haskell grammar to be able to process its programs. To achieve this, the following tasks needs to be performed.

- Analysis of the Haskell grammar
- Specification of the Haskell lexer
- Specification of the Haskell grammar syntax for parser

First, the analysis of Haskell grammar specification needs to be realized. Grammar specification we used [1] is expressed in EBNF form as HTML document. To be able to process the grammar programmaticaly, it is required to transform HTML form of grammar to more suitable representation. To transform the Haskell grammar to the form we need, we have created a grammar transformator. This tool first transforms HTML representation of Haskell grammar to XML representation which is more suitable for further processing. The advantage of XML is a tool support that allows to easily parse and transform XML file to the desired form. XML structure reflects the grammar rules and mapped particular grammar elements to XML elements.

XML grammar is then processed to create Java object model where XML elements are mapped to the instances of the grammar model classes. Java object model of Haskell grammar gives us a better way how to manipulate and operate on Haskell grammar. Class diagram of object model is in the fig. 2.

Mapping is shown in the example of a grammar rule `gdrhs ::= gd = exp ( gdrhs )?`. This rule is transformed to the following XML fragment.

```xml
<nonterm id="n1" label="gdrhs">
  <sequence id="seq1">
    <nonterm id="n2" label="gd" />
    <term id="equals" label="=" />
    <nonterm id="n3" label="exp" />
    <option id="opt1">
      <nonterm id="n1" label="gdrhs" />
    </option>
  </sequence>
</nonterm>
```
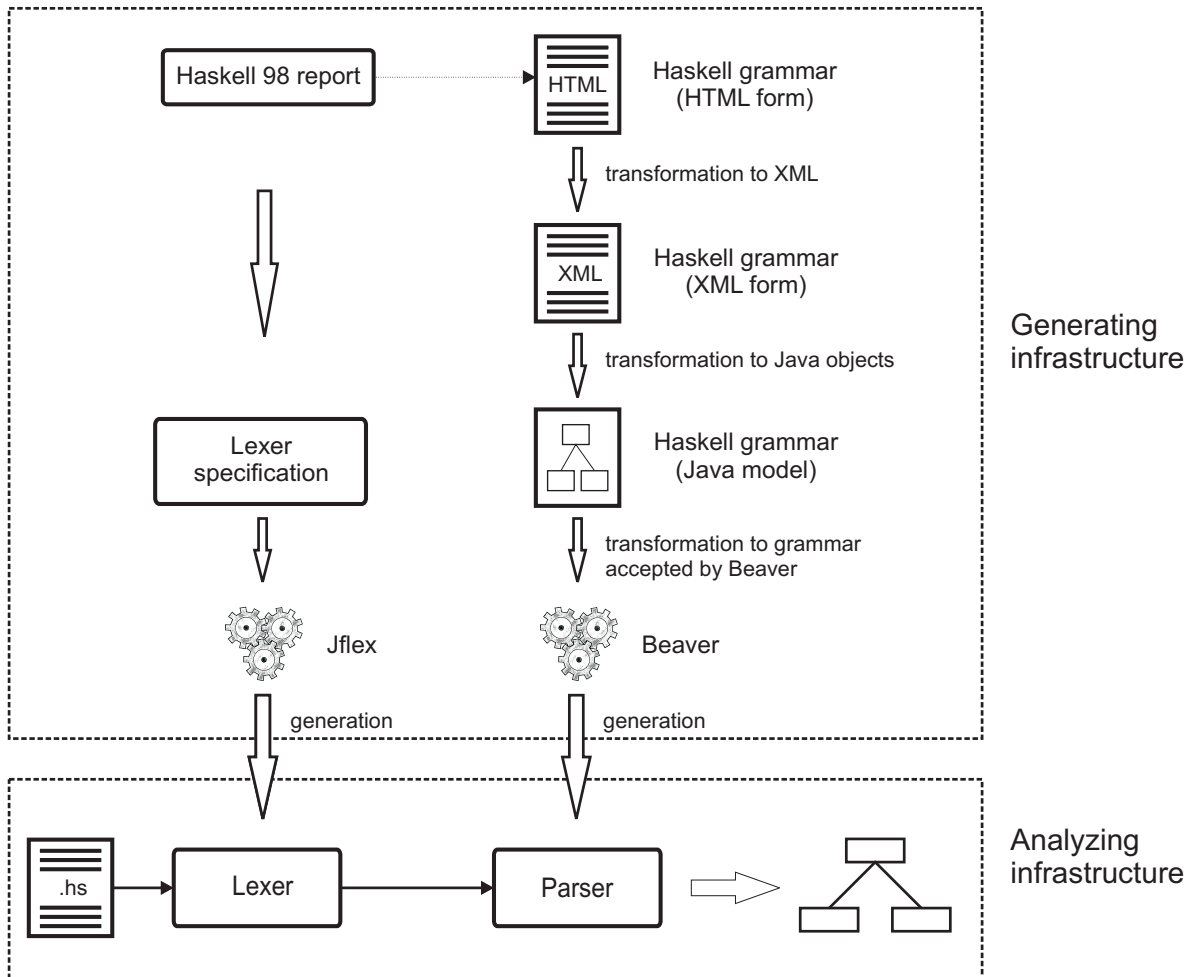
Fig. 1: Architecture of Haskell Syntax Analyzer tool

As we can see in [1], grammar specification consists of two parts - lexical syntax and context-free syntax. Lexical syntax and notes stated in [1] serves as a specification for the lexer usable in lexical analysis of Haskell programs. This specification, consisting of list of tokens and lexical syntax, is used as an input of JFlex [11] tool. Based on its input, JFlex generates lexer (scanner) of Haskell programs.

The second part of grammar specification represents context-free grammar of Haskell. This specification needs to be processed to help us to create a parser usable in syntactical analysis of Haskell programs. Based on analysis of the Haskell grammar [2] and form of needed input, we decided to use LALR parser generator Beaver [12]. Grammar transformator generates input EBNF grammar for Beaver parser generator based on grammar object model. Input for Beaver is extended by action routines that will be executed when parser reduces a group of symbols that define a symbol to the nonterminal. Action routines represent a code to generate derivation tree as a part of the translation process. Based on its input, Beaver parser generator generates parser of Haskell programs.

The described procedure and the used tools are the principle of functionality of the Haskell Syntax Analyzer generating infrastructure. Generation is realized with respect to Haskell grammar specification. Generated lexer (scanner) and parser are the tools of analyzing infrastructure and they are used to analyze Haskell programs.

[2]Grammar contains left recursion, therefore LL parser is not usable.
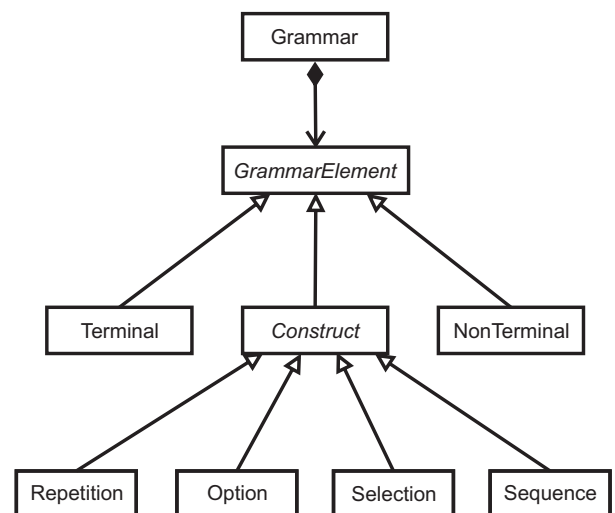
Fig. 2: Class diagram of Java object model of Haskell grammar

B. Analyzing infrastructure

Analyzing infrastructure of Haskell Syntax Analyzer is used to analyze Haskell programs. It consists of Haskell lexer and parser. The whole process of analysis works as following: Haskell program is scanned by lexer to produce the tokens

(lexems) used in the program. Produced tokens can have its attribute specified. Tokens are then used as an input for parser which produces the derivation tree.

To have a set of relevant data, Haskell Syntax Analyzer tool should be used with a large amount of programs divided to groups by their domain. This way we can analyze derivation trees of programs corresponding to particular domain and define the subset of Haskell grammar constructs which were actually used.

C. Example

Let assume that the Haskell program contains only one function:

```
a = [1,3..19]
```

Based on this program, generated Haskell lexer generates following tokens which are used in the program:

```
LBRACE
VARID<a>
EQUAL
LBRACK
INTEGER<1>
COMMA
INTEGER<3>
DOTDOT
INTEGER<19>
RBRACK
RBRACE
EOF
```

The list of used tokens is used as an input for generated parser. The parser processes tokens (LALR reduction according to the grammar) and constructs derivation tree of this program. Derivation tree is in the fig. 3.

Following the mentioned example, according to the Haskell Grammar [1], derivation tree is:

IV. CONCLUSION

Proposed Haskell Syntax Analyzer tool is used to analyze syntax of Haskell programs. The tool produces program derivation trees that can be further analyzed to get an insight about which Haskell grammar constructs are not used in a particular domain. Based on this analysis, we can reduce the set of grammar constructs and create a dialect of Haskell language optimized for the particular domain. It can be also extended by visualization of the analyzed data.

Architecture of the tool is designed in a way to be easily modified to process programs of other language. To process different language, specification of the language grammar should be provided. According to the new grammar, lexical specification and transformer to XML representation are modified. The rest of the architecture is reusable.

Deeper analysis of collected derivation trees can reveal patterns that are repeating in more programs. This patterns can be collected and reused by writing programs in the same domain.

ACKNOWLEDGMENT
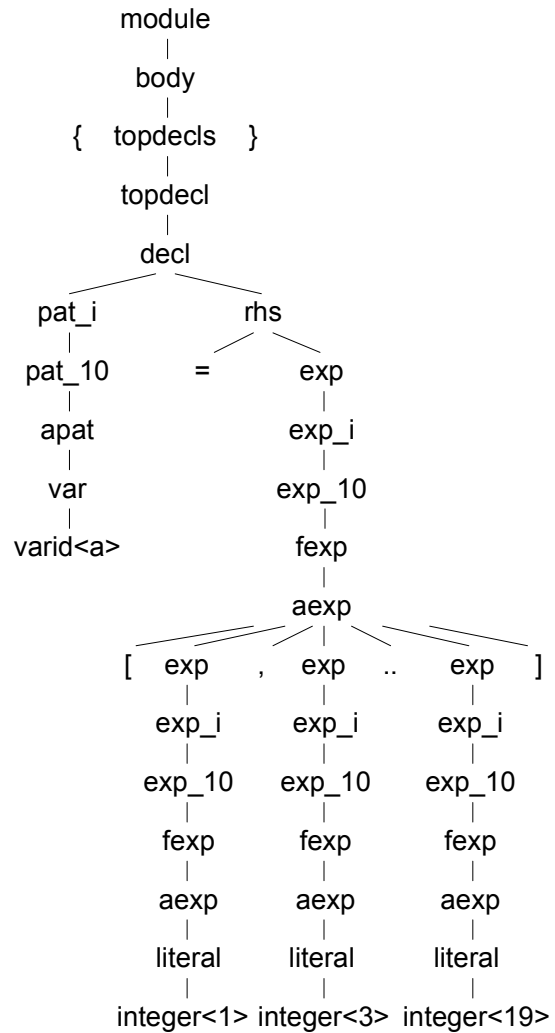
Fig. 3: Derivation Tree Example

REFERENCES

[1] S. P. Jones, "The haskell 98 language report," http://www.haskell.org/onlinereport/, 2002.

[2] A. van Deursen, P. Klint, and J. Visser, "Domain-specific languages: an annotated bibliography," *SIGPLAN Not.*, vol. 35, no. 6, pp. 26–36, 2000.

[3] M. Mernik, J. Heering, and A. M. Sloane, "When and how to develop domain-specific languages," *ACM Comput. Surv.*, vol. 37, no. 4, pp. 316–344, 2005.

[4] D. Spinellis, "Notable design patterns for domain specific languages," *Journal of Systems and Software*, vol. 56, no. 1, pp. 91–99, Feb. 2001. [Online]. Available: http://www.spinellis.gr/pubs/jrnl/2000-JSS-DSLPatterns/html/dslpat.html

[5] A. van Deursen and P. Klint, "Little languages: little maintenance?" *Journal of Software Maintenance*, vol. 10, no. 2, pp. 75–92, 1998.

[6] M. Fowler, *Domain-Specific Languages*. Addison-Wesley Professional, 2010.

[7] S. Dmitriev, "Language oriented programming: The next programming paradigm," http://www.onboard.jetbrains.com/articles/04/10/lop/, 2004.

[8] J. Greenfield, K. Short, S. Cook, and S. Kent, *Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools*. Wiley, August 2004.

[9] J. Gray, K. Fisher, C. Consel, G. Karsai, M. Mernik, and J.-P. Tolvanen, "Dsls: the good, the bad, and the ugly," in *OOPSLA Companion '08: Companion to the 23rd ACM SIGPLAN conference on Object-oriented programming systems languages and applications*. New York, NY, USA: ACM, 2008, pp. 791–794.

[10] J. Holmerin and B. Lisper, "Development of parallel algorithms in data field haskell," in *Euro-Par 2000 Parallel Processing*, ser. Lecture Notes in Computer Science, 2000, vol. 1900, pp. 762–766.

[11] G. Klein, "Jflex - the fast scanner generator for java," http://jflex.de/, 2009.

[12] A. Demenchuk, "Beaver - a lalr parser generator," http://beaver.sourceforge.net/, 2011.

Preliminary analysis of a european research organizations social network

Gabriel LUKÁČ

Department of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

gabriel.lukac@tuke.sk

Abstract—In this paper we provide preliminary results of an analysis of a social network of european research organizations participating on projects funded by European Commision. For our analysis we used a dataset of FP5 - FP7 Framework Programs provided by CORDIS. As a result of our preliminary analysis we give some basic statistical information about topological properties of the network, and we mention problems of ambiguous organization identification. In the end, future research direction is drafted.

Keywords—Social networks analysis, data mining, CORDIS, european research projects, framework programs.

I. INTRODUCTION

Research of social networks can bring new information about structure of various social entities and dynamical processes occuring between them. When we are talking about social network, we are not focused only on popular internet services like Facebook or Twitter that offer facilities for easy communication and sharing of resources between particular participants, but we will try to broaden our view to any network where some kind of binding between social entities occur. According to Wasserman et. al, by the term *social network* we will in general understand a finite set of actors and relations defined on them [1].

A more general definition of a social network permits us to think about a broader application impact. Research information diffusion models [2][3][4] try to describe a spread of influence [5] between organizations and businesses. For virologists and epidemiologists networks provide a framework to explore the spread of infections among human population [6]. A metaphor from epidemiological theory is being successfully used for marketing purposes [7].

II. RELATED WORK

Research of real-world networks has brought some interesting results regarding topological properties of networks and their dynamics. According to findings of [8], real-world network have completely different topological properties as historically popular theoretical models introduced by [9]. The substantial progress in discovery of interesting topological properties of real-world networks, several network models have been introduced [10] [11].

Collaboration networks have been extensively studied by several groups of authors. The structure of a collaborative network of authors participating together on writing a research paper has been studied in [12]. A similar topic to our work studied Barber et al. in [13]. They focused at detailed structural analysis of cooperation network of european research organizations participating on FP1 - FP4 framework programme projects.

III. SOCIAL NETWORK DATA

The data for the analysis of european research organizations have been provided by CORDIS [14]. It manages all of the information about projects and organizations participating in projects co-financed by European Commision. The dataset, that has been provided, covers data about projects realized in FP5 - FP7 Framework Programs. In order to keep the dataset actual as much as possible (the number of projects in CORDIS dataset is still growing as the Seventh Framework Program is still running), an automatic crawler has been implemented. It tracks changes on the CORDIS website and downloads new information about projects started after the date when the complete dump of CORDIS database was realized.

The dataset contains information about 35263 unique projects. There is 17628 projects from the Fifth Framework Program, 10660 from sixth and 6975 from seventh. From the projects data identifiers of participating organizations have been extracted and they have been used to build a social network graph. Regarding graph definition, organization identifiers represent vertices of the graph. There is an edge between particular couple of vertices, if these two organizations have participated on a same project.

Identification of respective organization in the dataset is very ambiguous since it could happen that the same organization participating in two or more projects could be identified in the dataset differently. Because of there is no strict rule how to create identifiers for organizations, it is hard to estimate how many different vertices actually represent the same organization. One the other hand, one could come up with a text-minig task to extract unambiguous organization identifiers from webpages of participants (for particular project), but according to same reasons (like absence of strict rules for creation of organization names participating in different projects), it is almost impossible task to do this task accurately. In a case of universities a typical ambiguous situation can be characterized by a faculty participating on a project and identified on a CORDIS webpage by its full faculty name, address and a city and on the other hand the same faculty is participating in another project with different partners and it is being listed on the CORDIS webpage with only an university name and the name of the city. One may think of clusterring organizations with same second names (like name of the university) within
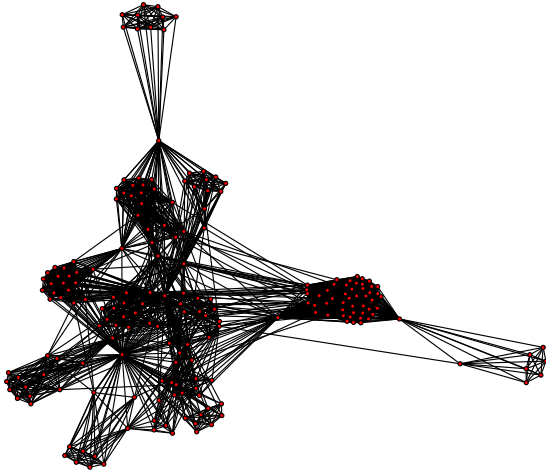
Fig. 1.   A sample of collaboration network where Technical University of Kosice has taken part (this is a sample of a filtered subset of original data where Tomas Sabol and Jan Paralic are listed as contact persons.

same cities and labeling them with a common identifier, but this kind of approach could bring a loose of granularity on the organization level and one will loose the opportunity to discover interesting relationships on the level of communities.

IV. RESEARCH ORGANIZATIONS COLLABORATION NETWORK PROPERTIES

Our social network was extracted from the dataset of projects and organizations participating on research projects funded by European Commision. More formally, the social network is represented by a weighted graph, where vertices represent organizations. There is an edge between two vertices if the two organizations (represented by the above-mentioned couple of vertices) have participated together on a same research project. The weight of an edge represent a number of projects, on which the two organizations have cooperated. On a Figure 1 one can see a sample of our collaboration network. The collaboration network visualized on the above-mentioned Figure shows collaborations in projects of Technical University of Kosice (we applied a filter to extract only those projects where Tomas Sabol (Faculty of economics, TU Kosice) and Jan Paralic (Faculty of electrical engineering and informatics) were listed as a contact persons).

A. Topological Characteristics of a Collaboration Network

In the following text we provide an overview of basic results(for details see Table I) of european research organizations collaboration network. Given results will drive further research and will enable us to compare the CORDIS network with other real-world complex networks.

In the un-preprocessed version of the network (for the preliminary results we haven't used any approach to merge two different identifiers of physically one organization) consist of 135236 vertices and 1381977 edges. Based on the overall number of vertices it is clear that visualising such a huge network is a non-trivial task. We have tried various open-source network visualization tools that are available on the web, but whatsoever efforts to visualize the network as a whole collapsed on memory resources of computer that we used. Regarding these problems we decided to select a potentionaly

TABLE I
OVERVIEW OF CORDIS NETWORK DATA

Measured Property	Result
Total number of vertices (framework programmes FP5 - FP7)	135236
Total number of edges	1381977
Number of vertices in FP5 graph	17628
Number of vertices in FP6 graph	10660
Number of vertices in FP7 graph (untill 30.5.2010)	6975
Number of connected components	4101
Number of vertices or largest connected component	107331
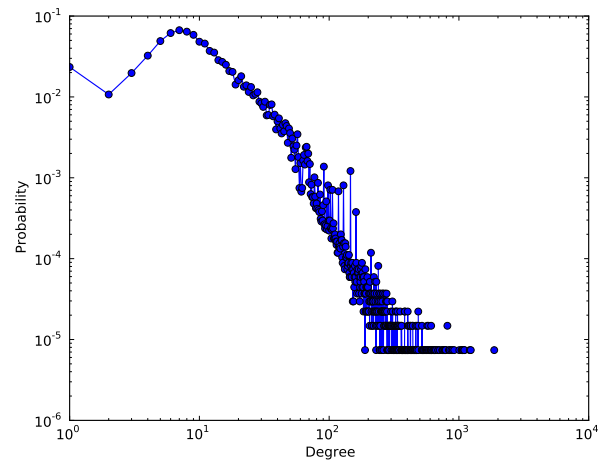
Fig. 2.   Degree distribution of the overall collaboration network.

interesting subset of a whole graph and visualized it (see Figure 1).

The degree distribution depicted on 2 could be treated as a function following the power law. According to the figure, it follows the power law with a small exception for organizations having the vertex degree smaller than 7. We think that this situation is caused by a constrain of social network representation meaning that organizations participating on a same project are composing a complete graph. The second reason for rising tendency of degree distribution function for degree values less than 7 is that the number 7 is the most frequent number of participants in consortium cooperating in research projects in the dataset.

The network consists of 4101 connected components. The largest connected components consists of 107331 vertices and 1338251 edges. The shape of its degree distribution function (see Figure 3) is almost identical to the overall degree distribution function. The second largest connected component in the row has only 157 vertices. The size of further connected components shortly descreases and there are thousands of trivial connected components (consisting of one or two vertices joing by an edge). We think that most probably this is caused by ambiguous identification of organizations within dataset and there is a chance that organization from these separated pairs of vertices are present in network with different identifiers.

V. CONCLUSION AND FUTURE WORK

In our paper we have made a preliminary overview of a structure of social network of european research organizations. We have mentioned some basic properties of the topology of
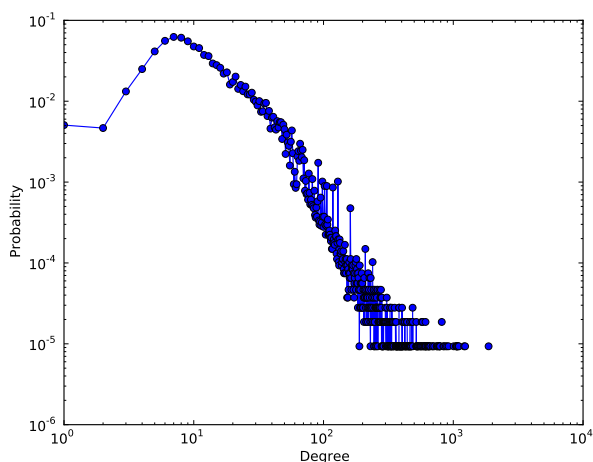
Fig. 3.   Degree distribution of the largest connected component.

the graph and we have pointed to a problem with ambiguous identification of particular organization in the dataset.

In the past, there has been an extensive effort in analysing the same domain as we are trying to do, but our work does not intent to reproduce them. Instead of that we plan to renew their results by observations on our data (Framework Programs FP5 - FP7). We would like to focus our attention on following directions:

a) Preprocessing: It is the most crucial task in order to get relevant results of network analysis. Regarding preprocessing, the ambiguous identifiers of same organizations should be unified. Althoughit will not be possible to guarantee a complete solution, we think that there is a possibility to achieve some level of precision. As a perspective approach to above-mentioned problem the work of Kostkova et al. seems to be promissing [15].

b) Data-mining: The second direction of our future research is going to be an area of data-mining. Here we want to focus on solving a network growth prediction task with the objective to predict which organizations will react to calls for proposals and possibly start new projects. We thinks that such an estimate could be helpful for funding agencies or ministries to better alocate funding resources for upcoming periods. The other possible future extension could be identification of scientific communities according to certain semantic annotations of projects they are involved in. Currently there are no such semantic annotations explicitly available in the dataset, but in general, there is an established practice to mention the research area of the project. It should be possible to use information extraction methods to extract such key text phrases and use them to annotate respective projects. We think that such an analysis will uncover typical patterns of collaboration between different areas of research.

REFERENCES

[1] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*, 1st ed.   Cambridge University Press, 1994.

[2] D. Kempe, J. Kleinberg, and E. Tardos, "Influential nodes in a diffusion model for social networks," in *Proceedings 32nd International Colloquium on Automata, Languages and Programming*, 2005.

[3] M. Granovetter, "Threshold models of collective behavior," *The American Journal of Sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.

[4] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst, "Cascading behavior in large blog graphs," Apr 2007.

[5] E. M. Rogers, *Diffusion of Innovations, 5th Edition*, original ed.   Free Press, August 2003.

[6] D. Easley and J. Kleinberg, *Networks, crowds, and markets : reasoning about a highly connected world*. Cambridge University Press, Jul. 2010.

[7] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," in *EC '06: Proceedings of the 7th ACM conference on Electronic commerce*.   New York, NY, USA: ACM, 2006, pp. 228–237.

[8] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, 2002.

[9] P. Erdös and A. Rényi, "On random graphs, i," *Publicationes Mathematicae (Debrecen)*, vol. 6, pp. 290–297, 1959.

[10] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science (New York, N.Y.)*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.

[11] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani, "Kronecker graphs: An approach to modeling networks," *Journal of Machine Learning Research*, vol. 11, pp. 985–1042, Feb. 2010.

[12] M. E. J. Newman, "The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 2, pp. 404–409, 2001.

[13] M. J. Barber, A. Krueger, T. Krueger, and T. Roediger-Schluga, "Network of european union–funded collaborative research and development projects," *Phys. Rev. E*, vol. 73, no. 3, p. 036132, 2006.

[14] P. Office. (2010) Eu-funded r&d projects website. [Online]. Available: http://cordis.europa.eu/

[15] K. Kostkova, M. Barla, and M. Bielikova, "Social relationships as a means for identifying anindividual in large information spaces," in *Artificial Intelligence in Theory and Practice III*, ser. IFIP Advances in Information and Communication Technology, M. Bramer, Ed.   Springer Boston, 2010, vol. 331, pp. 35–44.

Projection of Affiliation Network and Modifications of Positional Analysis

Martin REPKA, Gabriel TUTOKY

Dept. of Cybernetics and Artificial Intelligence FEI TU of Košice, Slovak Republic

{martin.repka, gabriel.tutoky}@tuke.sk

Abstract— **This article describes projection of two-mode networks onto one-mode network and approaches to positional analysis modifications. We discuss affiliation network projection background focusing on preparing suitable network data for positional analysis. Next we shortly describe some ideas of modifications using non-structural data in positional analysis. In the conclusion we mention future direction in our research.**

Keywords— **social networks, affiliation networks, positional analysis, structural analysis, social network data**

I. INTRODUCTION

In recent years, many "social networks" have been analyzed like Internet, email networks, peer-to-peer networks, telephone call graphs or train routers [1]. All of these networks are interesting in their basis and they are a notable data source for network analysis. There are usually large-scale networks with thousands of nodes and edges. Analysis of these networks, usually based on global properties, can lead to interesting and helpful results. Nevertheless, there exist many different situations in network analysis where data of these networks do not carry sufficient information.

Next, a more promising source of network data is affiliation network. An affiliation network is a network of actors connected by common membership in groups/events such as clubs, teams, organizations or others activities. Affiliation networks are special type of two-mode social networks [2] where one mode is a set of actors, and the second mode is a set of events which are affiliated to the actors. The tie between actors and events are created, if actor is member of a group/participates on particular event. Affiliation networks describe collections of actors rather than simply ties between pairs of actors and connections among members of one of the modes are based on linkages established through the second mode [2].

Affiliation networks were studied in past, e.g. studying attendance of women in social events [3], movies and their actors [4] or co-authorship network of scientists and their papers [1]. Whereas in the first two examples, the authors used unweighted representations of the networks, in the last work, the author introduced interesting approach for building of collaboration network where he used weighted ties between authors of the same paper. The weight of the tie between collaborated authors of a single paper is derived from count of the paper collaborators, and final weight of two collaborated authors is a sum of weights over all papers where authors collaborated. This approach allows finding the "most connected" scientists in the whole collaboration network.

In our work we build collaboration network of teenagers (described in next section) based on their participations on educative-pedagogic workshops for evaluating the "most connected" persons. In network building process we used Newman's approach for weighting of the ties between collaborators proposed in [1],[5].

II. DAK COLLABORATION NETWORK

DAK – community network is a collaboration network of members (usually teenagers) of a non-profit organization dealing with organizing educative-pedagogic workshops for young people [6]. Usually there are organized around 10 workshops annually with 150 – 700 participants on single workshop. The number of participants depends on workshop's type and duration. All participants of a single workshop are partitioned into smaller groups, usually with 8 – 12 members. Each group member cooperates with each other group member and so there are established new social relations between group members. Generally there are two types of group members – *group participants* and *leader(s) of a group*. Participants are "isolated" inside group (usually they do not establish social relations outside group), but leader(s) cooperate with other leaders and so they create another type of social relations. Additionally we recognize two types of groups – *participants group* and *organizers group*. In summary we have two types of groups and four types of group members: *base participant, base organizer, leader of participants' group* and *leader of organizers' group*.

Compositional attributes in DAK data set are available for both, actors and events. Actors are described by attributes such as date of birth (age) or gender; and by geographical attributes – city or area of living. Events are described by their type. We can recognize two main types of events – events for base participants and events for organizers. Events for organizers are next categorized by their types of activity like registration, security or accommodation event.

Moreover, temporal attributes are available together with compositional attributes, e.g. start and end of events/workshops. From these data we can derive several other attributes, such as "length of event" or "time of first visit" for particular actor. In our case it means the moment when the actor visited any event for the first time.

A. Two-Mode representation

Collaboration network described above can be expressed

for single workshop by a bipartite graph as representation of two-mode network (see Figure 2). The first mode is a set of actors, and the second mode is a set of events which affiliate the actors. Actors are all members of organization who participated at least in one event. Each group on the single workshop we represent as a single event, so for one workshop we obtain several events. Additionally we added two more events for representation of cooperation between leaders of participants and leaders of organizers.

One of the advantages of DAK data set is availability of temporal information in the data. We are able to track events in the time and recognize which events were organized in parallel and which sequentially. Also we are able to track participation of single actors on particular events.

B. Projection onto One-Mode representation

Affiliation networks (two-mode representation) are most often projected onto one-mode networks where actors are linked to one another by their affiliation with events (*co-membership* or *co-operation*), and at the same time events are linked by the actors who are their members (*overlapping* events). This property is called *duality* of the affiliation network [2].

Usually, weights in both, affiliation (two-mode) and also in projected (one-mode) networks have binary values. The ties in the networks exist or not. In the step of projection of two-mode networks onto one-mode networks we can use e.g. four different measures for weight definition by Opsahl [7]:

- Weight is determined as a count of participations (co-occurrences) – e.g. the count of events were two actors participated together, formalized expression is $w_{ji} = \sum_p 1$, where w_{ij} is the weight between actors (nodes of the first mode) i and j, and p are events (nodes of the second mode) where i and j participated on.

- Newman in [1],[5] proposed extended determination of weights while working with scientific collaboration networks. He supposes that strength of social bonds between collaborators is higher with lower count of collaborators on a paper and vice versa social bonds are lower with many collaborators on a paper. He proposed formula for defining the weights among collaborators as $w_{ji} = \sum_p \frac{1}{N_p - 1}$, where N_p is the count of collaborators on paper (event) p.

Till now we considered only binary two-mode networks and their projection to weighted one-mode networks. However, there exist also weighted two-mode networks, such as networks of online forums (weight is determined as count of posts or posted characters) or collaboration network described above and in [8, 9]. So, both just presented measures for weight definition could be extended for weighted two-mode networks as follows:

- $w_{ji} = \sum_p w_{i,p}$, where $w_{i,p}$ is the weight of ith actor to p^{th} event. This method differentiates how two particular actors interact with the common event, and to project it onto a directed weighted one-mode network. In this type of projected network, all dyads are composed of either two directed ties (mutual) or no directed ties (null) and no dyads are made of a single directed tie (asymmetric) [7].

- By a similar way, the Newman's method can be extended for projecting of two-mode networks. The function of weight determination is given by formula $w_{ji} = \sum_p \frac{w_{i,p}}{N_p - 1}$.

This formula would create a directed one-mode network in which the out-strength of a node is equal to the sum of the weights attached to the ties in the two-mode network that originated from that node [7].

III. ANALYSIS OF SOCIAL NETWORKS

A. Approaches

In social networks we can identify three ways of analysis. At first we can study global properties of a network as a whole. There can be analyzed several measures such as a measure of clusterization or an average degree of separation in network. These analyses provide basic knowledge about the network, mostly to support other analyses (in order to find or choose suitable algorithms for certain tasks in further analysis).

The second approach is aimed to study actors in social network. Those actors are considered as individuals and they are treated in this way. This kind of analysis computes and investigates individual properties of actors in the network such as various centrality measures (betweenness centrality, input degree) or connectivity and reachability of actors in the network.

The last and for us in this study the most interesting type of analyses is social groups' analysis in networks. The social groups in social networks can be identified in two ways:

- analysis of cohesive subgroups,
- analysis of positions and roles.

Main difference between analysis of cohesive groups and analysis of positions is that cohesive groups are identified through the reachability, common ties between actors or frequency of ties in group. So cohesive groups contain actors, who are relatively close to each other. Position groups actors which could be in various distances and it is rather dependant on the structure (substructures) in the network [10].

B. Positional analysis of networks

Positional analysis is based on groupings of actors based on defined equivalence. Actors who are equivalent form a class of equivalence and this corresponds to a position in positional analysis. The most common equivalences are structural, automorphic and regular equivalence [2].

Positional analysis is based on structural data of social networks, but in social networks exist other types of data such as compositional and temporal attributes. As we introduce local role equivalence (see next), we can propose some concepts of using this nonstructural data.

Local role equivalence is defined as equivalence between actors who are equivalent if they are in the same roles to other positions. Role is defined as a relation between two positions [2].

Using structural data we can execute original positional analysis. Compositional data are suitable for classification or clustering in dimensions created by compositional attributes. Temporal data are suitable for dynamic analysis such as network evolution and evolution of positions. Temporal attributes can also be used for network decomposition into time periods.

As we consider all mentioned kind of data, we can enrich and improve original positional analysis. We propose three

alternative ways how to use compositional data in positional analysis:

- dividing and joining of existing positions,
- fine tuning of positions,
- including in the definition of position.

Dividing and joining of existing positions is based on comparison of "similarity" of compositional attributes within and between positions. Fine tuning of positions is an approach, which solves cases of unclear decisions of actor's membership to position. For example if actor *A* is suggested into position *P1* for 40%, to *P2* for 40% and to other positions for 20%. It compares the distance based on compositional values from members of each position and decide to which position it inclines more. In the last case of including of compositional attributes in position definition the definition contains structural definition and some rules from compositional data [11].

C. Splitting and joining of positions
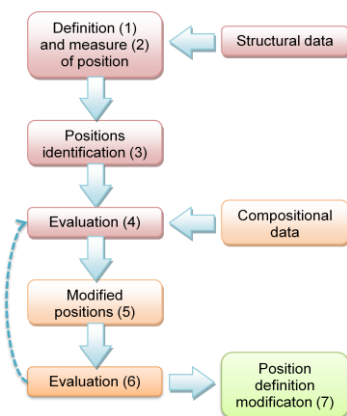
Figure 1 Process of modified positional analysis using splitting and joining of positions

This way of modification of positional analysis is only is very similar to the original method. It consists of seven steps. First we search positions using classical positional analysis and structural data, we might use local role equivalence. We define definition of local role equivalence (1) and then we proceed with definition of measure of local role equivalence (2). After that we proceed to the positional analysis (3) and evaluate obtained positions (4). So far we described the standard process of positional analysis, now our modification comes into the play. After evaluation of positions (4) we make corrections to the existing positions in comparative way (we compare similarity of compositional attributes between position and within position) (5). It can result in two alternative decissions:

1. within one position several clusters are formed due to information from compositional attributes,
2. actors in some positions could be very close due to compositional attributes (form a cluster) amongst all actors in network.

In case 1 these actors in one position could be split in several positions or as in case 2 the actors in different positions could be joined to form a single position. The modified positions are evaluated (6) and if the modification is resolved as helpful the definition of equivalence of actors could be changed appropriately (7).

On figure 3 we can see an example of community network, which we analyze for position using local role equvalencie. From the structure we obtained basic diagram of positions as can be seen on figure 4 (1). If we extend the diagram of positions (2)
we can see some substructures, which can form other positions. We can examine compositional attributes and

decide where the positions could be split or joined. At the bottom of the picture we can see possible configuration of positions after process of splitting and joining of positions.
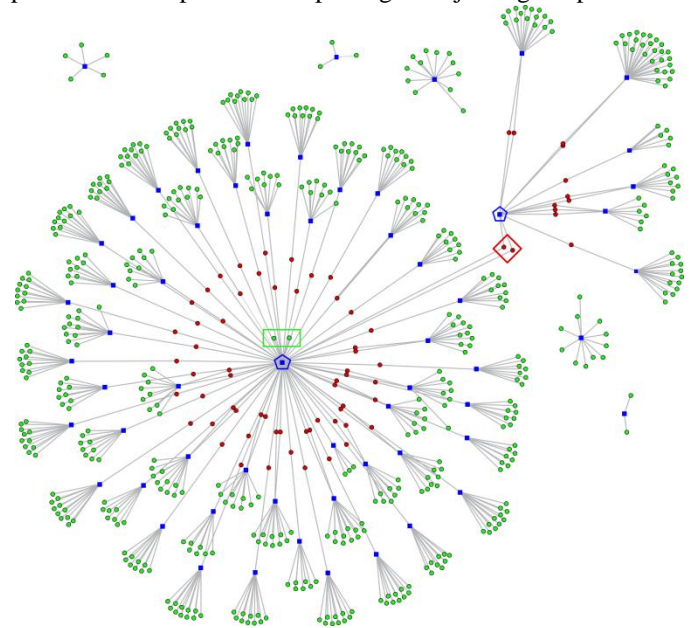
Figure 2 Example of two-mode social network - community network
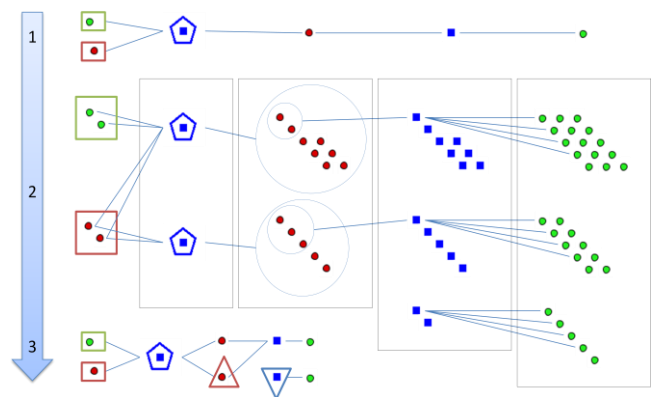
Figure 3 Modifying the positions by splitting or joining

D. Fine tuning of positions

Fine tuning is another modification of positional analysis, where compositional attributes are used in case of uncertainty when resolving a membership of actor to the position (2). First step is to define a definition of equivalence based on the structure (1). Then we need to define a measure of equivalence using structural data and for uncertain

Figure 4 Process of modified positional analysis using fine tuning of positions

situations also compositional data (2). Now we can resolve positions (3) and evaluate them (4). From results of evaluation we can decide to do some changes in definition of equivalence (5).

E. Including in the definition of position

The last option of utilization of compositional data is to
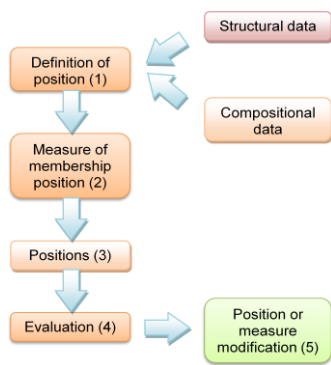
Figure 5 Process of modified positional analysis including compositional data in definition

use them directly in the definition of equivalence. So the definition of position is defined based on structural and compositional data together (1). Measure of membership to position is also defined in structural and compositional attributes (2). Then the positions are identified (3) and evaluated (4) and corrections to definition are made (5). The main difference between previous principles is the step in the process when the compositional data are introduced. In this way the compositional data are included at the beginning of the positional analysis.

IV. CONCLUSIONS AND FUTURE WORK

In our work we described and built collaboration network of teenagers (described in section II) based on their participations on educative-pedagogic workshops. We evaluated network as whole. In the future work we will evaluate this network in desired time stamps. It means, that in the network building process we will use approach for weighting of the ties between collaborators based on weighting used by Newman in [1, 5] and time/frequently dependent weights decreasing over the time/workshop count. We assume that weight of two collaborators is dependent on the count of event participants and together on the time between events where these actors collaborate. Also we assume that weight decreases with increasing number of event participants and also with longer spare time /less frequency between events.

In the beginning of section III we introduced some approaches of structural analyses in social networks especially positional analysis and we proposed three ways of modification of positional analysis using nonstructural data. Considering nature of positional analysis with compositional data included in the process it has big potential to improve accuracy and adequacy of identified positions. These hypotheses should be proven in experiments on several various two-mode social networks with verification by experts or administrators of particular social networks. Affiliation networks, which is a special type of two-mode networks are very suitable for positional analysis and usually contains quality compositional and temporal data. As target of our social network analysis we will target mainly on community networks (DAK), bibliographic networks and business social networks.

ACKNOWLEDGMENT

REFERENCES

[1] M. E. J. Newman, "Who is the best connected scientist? A study of scientific coauthorship networks," *Complex Networks*, p. 32, 2004.

[2] S. Wasserman and K. Faust, *Social Network Analysis*. Cambridge: Cambridge University Press, 1994.

[3] A. Davis, *et al*., *Deep South*: University of Chicago Press, 1941.

[4] D. J. Watts and S. H. Strogats, "Collective dynamics of 'small-world' networks," *Nature*, p. 3, 1998.

[5] M. E. J. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," *The Americal Physical Society*, vol. 64, p. 7, 2001.

[6] *DAK - Collaboration Network, data set of non-profit organization*. 2011. Available: http://www.domcek.org

[7] T. Opsahl, "Projecting two-mode networks onto weighted one-mode networks," vol. 2011, ed, 2009.

[8] G. Tutoky and J. Paralič, "Modelovanie a analýza malej komunitnej sociálnej siete," in *5th Workshop on Intelligent and Knowledge oriented Techniloties*, Bratislava, 2010, p. 4.

[9] G. Tutoky, et al., "Structural analysis of social groups in collaboration network," in *Faculty of Electrical Engineering and Informatics of the Technical University of Košice*, Košice, 2010, p. 6.

[10] M. Repka, "Analýza určitých typov sociálnych sietí". Písomná práca k dizertačnej skúške. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2011. 75 s.

[11] M. Repka, "Analýza vybraných typov sociálnych sietí". Ústna prezentácia k dizertačnej skúške (Powerpoint prezentácia). Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2011.

PROOF CLUSTER FOR ALICE ANALYSI FACILITY

Lucia Vaľová

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

lucia.valova@tuke.sk

Abstract— ALICE (A Large Ion Collider Experiment) at the LHC plans to use a PROOF cluster at CERN an another places for analysis. One of the cluster is the JRAF (JINR Russia Analysi Facility). The system is expecially aimed at the prototyping phase of analyses that need a high number of development iterations and thus require a short response time. Typical examples are the tuning of cuts during the development of an analysis as well as calibration and alignment. Furthermore, the use of an interactive system with very fast response will allow ALICE to extract physics observables out of first data quickly [1].

An additional use case is fast event simulation and reconstruction. A test setup consisting of Simulated and measured data will be located on local disks of workers. It can be used to perform analysis and calibration. The aim of the JRAF is conceptionally different from analysis on the Grid. The JRAF will not make it possible to analyse all data taken by ALICE because its space is limited. However, it is possible to run an analysis and see results after a few minutes or even seconds, thus allowing very fast development cycles.

Keywords—PROOF, ROOT, ALICE, JRAF, Grid, datasets

I. Introduction

Today's high energy physics experiments, like ALICE (one of the four bigger experiments on LHC at CERN), produce tremendous amounts of data (in the order of several PB per year). Because such an amount of data cannot be processed on single batch farms Grid solutions have been developed. ALICE's software runs in batch mode on the Grid. It performs large-scale production or reconstruction as well as analysis over large amounts of data. This model introduces a delay in obtaining the results. Users receive them after a given time which is the execution time of the program itself plus an overhead that tends to be bigger than in case of batch systems. Some tasks have a short execution time and need many iterations. Examples are the development of algorithms and analysis code as well as calibration and alignment procedures. The use of a local computer (which has clearly the lowest overhead) is feasible only if the data volume to be processed and CPU time needed are reasonably small.In addition to the Grid solution, ALICE decided to offer to its users a system that allows to process a big amount of data (compared with a single machine) without the need of submitting a Grid job [1].

The Parallel ROOT Facility, PROOF, is an extension of ROOT enabling interactive analysis of large sets of ROOT files in parallel on clusters of computers or many-core machines. More generally PROOF can parallelize the class of tasks the solution of which can be formulated as a set of independent sub-tasks (*embarrassingly* or *ideally* parallel).

The *main design goals* for the PROOF system are:
- **Transparency**: there should be as little difference as possible between a local ROOT based analysis session and a remote parallel PROOF session. Typically analysis macros should work unchanged.
- **Scalability**: the basic architecture should not put any implicit limitations on the number of computers that can be used in parallel.
- **Adaptability**: the system should be able to adapt itself to variations in the remote environment (changing load on the cluster nodes, network interruptions, etc.) [2].

II. The PROOF system

PROOF (Parallel ROOT Facility) allows interactive parallel analysis on a local cluster. Interactive means that you see the results right away (contrary to a batch job where you have to wait for the job to finish before you see the results). Parallel means that several nodes execute sub sets of your data at the same time.

PROOF is primarily meant as an interactive alternative to batch systems for Central Analysis Facilities and departmental workgroups. However, thanks to a multi-tier architecture allowing multiple levels of masters, it can be easily adapted to wide range virtual clusters distributed over geographically separated domains and heterogeneous machines (GRIDs).

While pure interactivity might not always be possible when performing a complicated analysis on a very large data set, PROOF still tries to give the user the interactive experience with something we call "interactive batch". With "interactive batch" the user can start very long running queries, disconnect the client and at any time, any location and from any computer reconnect to the query to monitor its progress or retrieve the results. This feature gives it a distinct advantage over purely batch based solutions, that only provide an answer once all sub-jobs have been finished [3].

You connect to a PROOF system from your usual ROOT prompt. Using PROOF is aimed to be transparent, that means you can execute the same analysis code locally and on a

PROOF system. Of course, certain rules have to be followed. A schema of a PROOF system can be seen here:
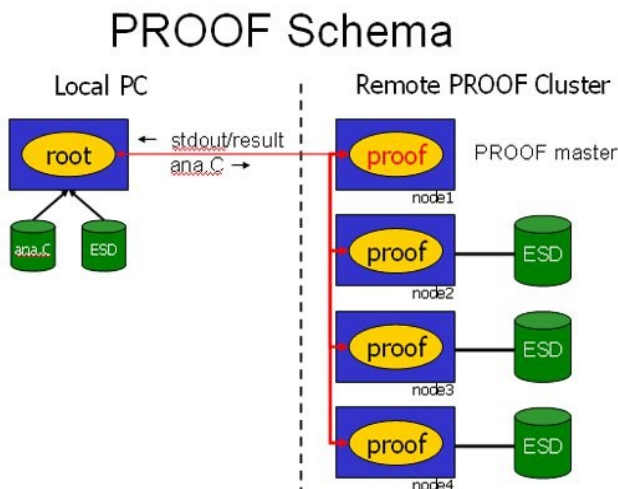
Fig. 1. PROOF Schema

PROOF is primarily meant as an alternative to batch systems for Central Analysis Facilities and departmental work groups (Tier-2's and Tier-3's) in particle physics experiments. However, thanks to a multi-tier architecture allowing multiple levels of masters, it can be easily adapted to a wide range of virtual clusters distributed over geographically separated domains and heterogeneous machines (GRID's).

The PROOF technology has also proven to be quite efficient in exploiting all the CPU's provided by many-core processors. A dedicated version of PROOF, **PROOF-Lite**, provides an out-of-the-box solution to take full advantage of the additional cores available in today desktops or laptops.

Apart from the pure interactive mode, PROOF has also an **interactive-batch** mode. With *interactive-batch* the user can start very long running queries, disconnect the client and at any time, any location and from any computer reconnect to the query to monitor its progress or retrieve the results. This feature gives it a distinct advantage over purely batch based solutions, that only provide an answer once all sub-jobs have been finished and merged.

The ALICE is one of the four bigger LHC experiments at CERN, Geneva. LHC is a particle accelerator with a circumference of 27 km. ALICE will study strongly interacting matter and the transition to the quark-gluon plasma. The Grid is intensively used for MC simulation and reconstruction.

The new type ALICE Analysis Facility (AAF) running PROOF will be used for :
 – Prompt analysis of proton-proton data
 – Pilot analysis of PbPb data
 – Calibration & Alignment.

Where fast response time is the key. The ALICE Analysis Facility is a distributed PROOF cluster used for interactive parallel data processing. There is combined ROOT`s package PROOF, with settings XROOTD (ALICE SE), which is

responsible for working with data, where the Packman ensure the timeliness software across a cluster.

III. THE CERN ANALYSIS FACILITY (CAF)

The CAF (CERN Analysi Facility) will run PROOF for ALICE. Simulated data and measured data are available on the CAF. It can be used to perform analysis and calibration. The aim of the CAF is conceptionally different from analysis on the Grid. The CAF will not make it possible to analyse all data taken by ALICE because its space is limited. However, it is possible to run an analysis and see results after a few minutes or even seconds, thus allowing very fast development cycles.

The design goal for the CAF is a system with 500 CPUs. At least 50 TB of selected data will be available. The CAF will also interface with the CERN Castor 2 disk cache to retrieve data selected for analysis. [5].

A schema of the CAF can be seen here:
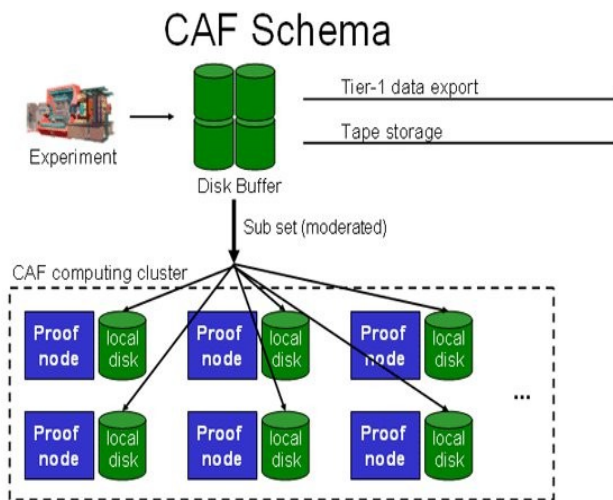
Fig. 2. CAF Schema

CAF user information

Detailed information on ALICE Analysis Facilities (AAF) and CAF is available at the www.aaf.cern.ch.

IV. RESULTS

This new type of PROOF cluster was setup at JINR (Joint Institute for Nuclear Research) in Dubna and is called JRAF (JINR Russia Analysis Facilities), see Table 1 [4].

			Cluster			ROOT	Aggregated disk space		
Name	Online	Status	Proof master	Workers	Users	Version	Total	Free	Used
1. CAF		Stable	alice-caf.cern.ch	82	1	v5-28-00a	119.2 TB	37.92 TB	81.32 TB
2. JRAF		Maintenance sin...	jraf.jinr.ru	8	0	v5-28-00a	2.014 TB	1.858 TB	160 GB
3. KIAF		Maintenance sin...	kiaf.sdfarm.kr	48	0	v5-27-06c	798.2 GB	199.1 GB	599.1 GB
4. SAF		Maintenance sin...	nansafmaster.in2p3.fr	48	1	v5-28-00a	12.07 TB	2.565 TB	9.507 TB
5. SKAF		Stable	skaf.saske.sk	60	1	v5-28-00a	53.72 TB	10.02 TB	43.7 TB
6. SKAF_TEST		Testing	skaf-test.saske.sk	2	0	v5-28-00a	815.9 GB	676.7 GB	139.2 GB
7. TAF		Open to local u...	pmaster.to.infn.it	16	0	v5-28-00a	3.914 TB	1.704 TB	2.21 TB
Total				264	3		192.5 TB	54.92 TB	137.6 TB

Table 1: Cluster list

Aggregated network traffic at this PROOF cluster is presented on Figure 3.
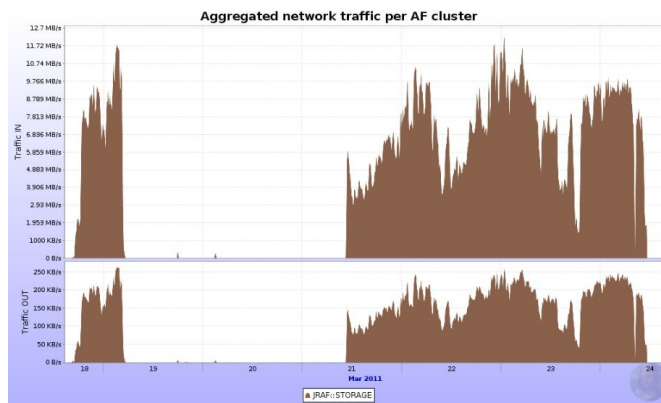
Fig. 3. Aggregated network traffic at JRAF

In order to analyze data, it is necessary to store it on the PROOF cluster. In the case of the AAF the storage is directly on a local drive of each computer in the cluster. It was necessary to ensure to copy data from Alien to the cluster catalog. To accomplish this task it was necessary to create a so-called dataset, which is basically a list of files. Data are divided into two groups:

a) official (real data, data from Monte-Carlo simulation),

b) user datasets (datasets created by users).

One of the advantages of PROOF is, that the user can list all of the active/waiting/finished sessions. If a cluster is overloaded the user can connect to another one. If the PROOF cluster has more than a specific number of users, the PROOF master will reject the new requests. The user can connect later and choose another AAF proof cluster. The system admin can kill/suspend the sessions in the queue. He can also change the position of the sessions.

V. Conclusion

I setup AAF (ALICE ANALYSIS FACILITIES) on 8 core computers in JINR (since I was there for a study visit).
The JRAF is one of only six AAF clusters. It was setup after CAF and SKAF. So we were one of the first for setup a cluster.
I monitor and I upgrade the AAF, when it is needed. This I know from the web pages aaf.cern.ch or from the TF (Task force) meeting and AAF meetings, which are every thursday.
I created datasets (a dataset is list of the files in one run). You can access list of datasets from PROOF by the command: "gProof->ShowDatasets();"

Monitoring of data transfer of JRAF can be found on the web page: http://alimonitor.cern.ch/display?page=PROOF/SEs
and Monitoring of JRAF status can be found on the web page: http://alimonitor.cern.ch/stats?page=PROOF/list
During the instalation I helped to find the bugs in AAF software and I report them directly to developers or in web page: https://savannah.cern.ch/projects/aaf/ [6-8]

References

[1] J. F. Grosse-Oetringhaus, The CERN Analysis Facility – A PROOF Cluster for Prompt Physics Analysis, CERN PH/ALICE
[2] www.aaf.cern.ch
[3] J. F. Grosse-Oetringhaus,The CERN Analysis Facility – A PROOF Cluster for Day-One Physics Analysis
[4] A . Bogdanov, L. Jancurova, A. Kiryanov, V. Kotlyar, V. Mitsyn, Y. Lyublev, E. Ryabinkin, G. Shabratova, S. Smirnov, L. Stepanova , W. Urazmetov, A. Zarochentsev, Distributed Russian Tier-2 – RDIG in Simulation and Analysis of Alice Data From LHC, J.Phys.Conf.Ser.219:072054,2010
[5] http://root.cern.ch/drupal/content/proof
[6] http://aliweb.cern.ch/Offline/Activities/Analysis/CAF/index.html
[7] http://alimonitor.cern.ch/stats?page=PROOF/list
[8] https://savannah.cern.ch/projects/aaf/

Proposal control of manipulators controlled via technological network DeviceNet.

[1]Matej ČOPÍK, [2]Ján Ilkovič

[1]Department of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic
[2]Department of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]matej.copik@tuke.sk, [2]jan.ilkovic@tuke.sk

Abstract — this paper deals with the proposal for the control of manipulators whose drives are connected to the technological network DeviceNet. These manipulators are integrated in the school model of the flexible manufacturing system. The main objective is to explain the employment of technological networks in the control and implement this knowledge in teaching process.

Keywords — Flexible manufacturing system, DeviceNet, Programmable Logic Controller, Manipulator, Drive

I. INTRODUCTION

In practice, often used interconnection of individual parts of the production through communication networks, because controlled processes are often large and use the network is cheaper and easier. It is the same for manipulators that controlling deals this paper. Manipulators whose controlling will address this paper are integrated in a flexible manufacturing system (FMS), which was created for the purpose of teaching at the Department of Cybernetics and Artificial Intelligence at the Technical University in Košice. This model project was funded by the Centre of Excellence. FMS shown in Fig. 1 is a production system, which includes machinery and equipment. In which are integrated sensors and actuators, some of them communicate using technology network with the Programmable Logic Controller (PLC). One of such network is the technological network DeviceNet. Through this network are connected to each drive manipulator in FMS. FMS consists of two tri-axial manipulators whose two axes are controlled by drives attached to the DeviceNet network. For more detailed description of this model, see [1] and [2].

The first part briefly describes the activities of the FMP which are integrated tri-axial manipulators. This part contains a description of the manipulators drives connected to the DeviceNet network.

The second part deals with the proposal control and configuration technological network DeviceNet, which is connected to the control PLC.

The third section describes the actual proposal of manipulator drive control.

Fig. 1. Model flexible manufacturing system

II. DESCRIPTION OF CONTROL PROBLEMS

A. Description FMS and manipulators

Flexible manufacturing system consists of six posts, which are carried out each step of the manufacturing process. The final product is a color image stack in template.

The first post is post implementation templates. This post is a first manipulator with indication MAXP12R-H41BR-C41BR0300. The selected color squares from the small cube trays and store them according to the. The model is entered from the information system (IS), into the template of the size of 5x5 blocks. Movement in X and Z axis provide drives, which have three-phase stepper motors, gearboxes and integrated electronics. These drivers are controlled via technology network DeviceNet. Movement in the Y axis is performed with a pneumatic piston drive.

On the second post is camera control which checks the accuracy stacked picture. After camera control the template for this position moves from one conveyor to the second conveyor.

On the third post is the second tri-axial manipulator with indication MAXR12R-S41BR1000-P41BR0600, which saves the template to rack or in their withdrawal from collecting from the rack. To rack can store up to 28 templates. Drives moving in X and Y axes. The drivers include the three-phase stepper motors, gearboxes and electronics. Drive in the Y axis is equipped with a brake that arm does not fall into the down position when power failure. Also, these drives are connected via technology network DeviceNet. Movement in the Z axis carries pneumatic piston.

On the fourth post is only the transfer between second conveyor and first conveyor.

On the fifth post the templates that are intended for removal, or the wrong stacking, emptied. Templates are emptied by the tipping manipulator, which tilts the template over the cubes container and clear the template.

Sixth post, post sorting cubes are located outside the production cycle. Cubes come from the vibrating container and are sorted into individual small storage cubes, where are then using a manipulator in the position one stored into the template.

Deployment of posts FMS is shown in Fig.2.
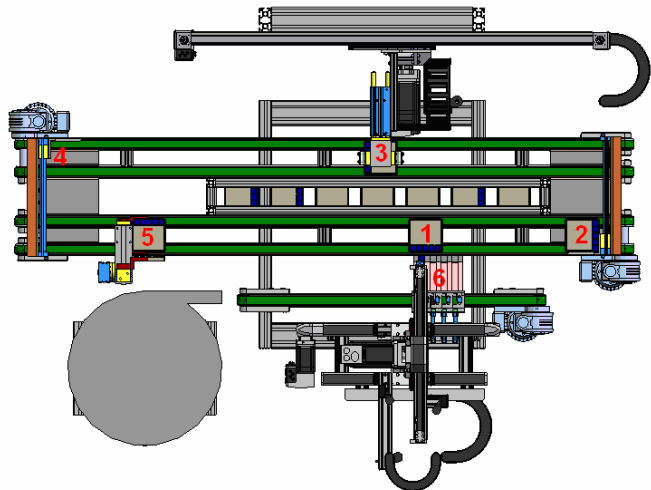
Fig. 2. Deployment of individual posts flexible manufacturing system

B. Description of the connection drive manipulators

Individual drives in both the manipulators in terms of control, administrative autonomy, even if they perform parallel operations. Each drive on the DeviceNet network is a one node of network, which communicates via a master module with control PLC. Master module is a scanning module with the type designation 1769-SDN, which is located in the PLC as an expansion card. This arrangement of drives on DeviceNet network is shown in Fig. 3 [3].
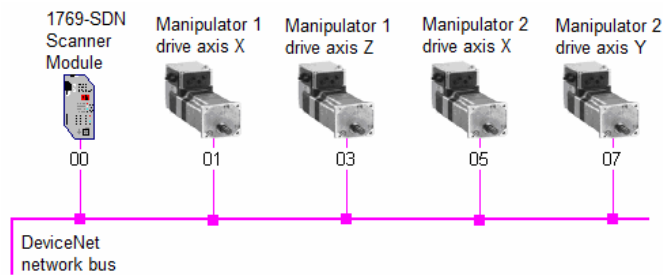
Fig. 3. Connection drivers to the network DeviceNet

III. CONTROL PROPOSAL

A. DeviceNet network configuration

Network nodes, thus particular drives, can have arbitrary addresses. The first node of network is DeviceNet scanner module, which also has the first address on the network - 00. Particular network nodes can be numbered from 00 to 63, so it constitutes a possibility to connect 64 nodes to one network. Chosen addresses of particular drives nodes are mentioned in table 1.

TABLE I
TABLE OF ADDRESSES FOR NETWORK NODES

Node number	Description
01	Motor of axis X for manipulator 1
03	Motor of axis Y for manipulator 1
05	Motor of axis X for manipulator 2
07	Motor of axis Y for manipulator 2

For possibility to add new nodes to the network in cases when PVS was modified addresses were defined in such a way as it is shown in Tab. 1. It is also possible to add a new network node in such a way that it is clear which nodes belong together with this numbering. For instance, if it was needed to add a new node into the network on the first post, thus for manipulator 1, it would be possible to use address 02 or 04. All nodes of the first post would be side by side in this case.

Also, it is possible to make network nodes reconfiguration by assignment addresses from already connected nodes. However, this procedure is rather difficult, because it is necessary to set addresses for particular nodes in hardware profile. Concretely for manipulators motors, it can be done by two rotary switches S1 and S2 on each motor. A decimal number of nodes address can be set by switch S1 (MSD) and a unit number of nodes address can be set by switch S2 (LSD). Switches for setting addresses of motor are depicted in Fig. 4. More detailed settings are listed in [4].

Fig. 4. Setting addresses of motors

B. Basic functions for motors control

Additional functions, which are directly created for development environment RSLogix 5000 by Rockwell Automation Company, are used for this kind of motors control by DeviceNet network. Overall there are 24 created functions, which carry out different functions from single data sending and receiving from and to DeviceNet network (*MC_UpdateInputData_DVN_ILx2D,*

MC_UpdateOutputData_DVN_ILx2D), through functions for motors control (*MC_Power_DVN_ILx2D, MC_MoveAbsolute_DVN_ILx2D*), to function for motors diagnostics (*MC_ReadStatus_DVN_ILx2D, MC_ReadParameter_DVN_ILx2D*). All functions for control of drives are listed in [4]. The function MC_*UpdateInputData_DVN_ILx2D* must be carried out as first at control cycle. This function provides messages acquisition, which control manipulators motors from DeviceNet network. Message sending to network is carried out by function *MC_UpdateOutputData_DVN_ILx2D*. It is necessary to put this function as last from added functions at the cycle. Different functions for control or diagnostics can be added between these functions. Every motor has accurately defined an array of variables, which it uses for communication and control. It is necessary to use data structures appertaining to particular motor for its control. All motors have included control electronics at themselves, which makes possible communication by DeviceNet network. It also provides direct connection of some signals (for example: from sensors at and positions).

C. Proposal for the control

The main problem in controlling these types of drives is that when a power failure can not remember your last position. Therefore it is necessary each time you start control program make the initialization at zero or the desired position. In this case, the initialization is made so that both drives the manipulators pass in its extreme positions, where they stop at the end position sensors from that position pass into the desired position. After reaching the desired position, this position set as the default for other movements, to what serves the function *MC_SetPosition_DVN_ILx2D*. After completion of the initialization it can be included in the production process. To initialize the drive position is also possible to use already created function *MC_Home_DVN_ILx2D*, but to use this feature, you must have an end position sensors attached directly to the control unit to drives.

To carry out operations themselves is the creation of appropriate control sequence, which is in the program call in any control cycle and at the beginning of the cycle only changes the parameters of position, since movements drives are made with regard to the starting position.

D. Proposal for the control of manipulator 2

The first manipulator that stores blocks of small storage cubes to the template on the conveyor is shown in Fig. 5.
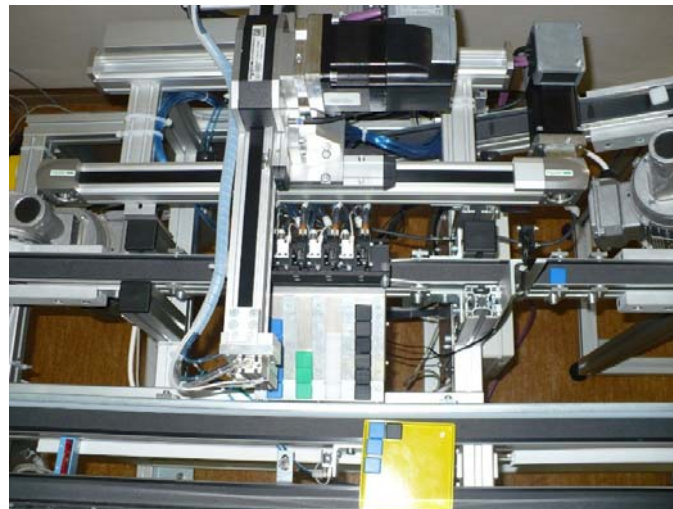
Fig. 5. First manipulator

Information on color picture, which has to be stack in template, is given a request for stack of the information system. Upon arrival requirements for stack resets the counter cubes. Subsequently, progressively from the requirements for stack gradually fetched serial numbers in a palette of cubes with their corresponding color. Progressive loading and storage cubes are made on the basis of two counters. The first is counter selected blocks, to be storage and the second counter is stored cubes those already stored on the template. If the values of these counters are the same, it is a sign of that is to grab a cube from the tray. In that case, the first color is loaded cubes as required. Based on the color of the loaded dice will send the drive in the X axis value on which to move in order to grasp the cube colors. Drive in the Y axis is sent in this case always the same value independent of the color cube, since all the containers have the same position in the axis Y. If the cube is grasped in the manipulator, the counter increments the selected blocks. A different value of the counters is a sign that has a cube that is stored in templates. Counter to the value stored blocks are sent driven in X and Y axis values, which corresponding to the location in the template. After saving the cube into the template increments the counter stored cubes. Cycle of grab and storage cubes is repeated, but with other values, which correspond to the next cube. After stacking all of the cubes changed template status to stack and is sent to the next post.

E. Proposal for the control of manipulator 2

The second manipulator within the production mode takes out empty and compiles full patterns into the rack during production. On the contrary, within discharge mode it takes out full and compiles empty patterns. The second manipulator is shown in Fig.6.

Fig. 6. Second manipulator

During production of new patterns the function for browsing the store is called and it determines where in the store an empty template is located. Based on the position in the store, values to the manipulator drives are sent. After transfer to a given position a template is taken and the manipulator moves it to a precisely defined place on the conveyor. While manipulator is moving towards the conveyor, same positions are sent with the drive. When a production process is completed and the template is supposed to be put in store, the opposite procedure begins. Function for searching an empty place in the store is called and at the same time manipulator moves to a fixed defined position where it takes a template from the conveyor. After taking the template based on a found empty position in the store, appropriate values are sent to drives. These values represent empty place in the store. After putting the full template in the store, this place is assigned a number that corresponds with the bar code of manufactured template.

For discharging, the control has a similar procedure where cubes are emptied from the full template in the store into the vibrating container. Based on the bar code, the template that should be discharged, template position is determined. This position is sent to manipulator drives that are transferred to a given position within a store. It takes a template in given position a puts it on a defined place on the conveyor. It is necessary to place an empty template in the store after discharging. Function for searching the empty place in the store is called and at the same time is the template taken. Values are sent to manipulator drives that correspond with empty place in the store. Subsequently, the template is put in the empty place in the store.

Functions for searching an empty place are called only in cases when we do not know to determine the position otherwise. If a full template is in the store and has a bar code assigned, position in the store can be determined.

Each position in the store corresponds with drives to which corresponding drives should move so that they are at a desired position. These values are kept in the chart and it is from here that corresponding values are sent to drives according to the position in the rack. For example, positions 1 to 7 has the same position in the axis X and only axis Y changes. Similarly, positions 1,8,15 and 21 have the same axis Y and only the axis X changes. It means the corresponding position is made up of seven positions in axis X and four in axis Y according to desired position in the rack.

IV. Conclusion

Realized control does not have to be finished. Further requirements may occur with time and they will have to be implemented in the control process. Various functions for diagnostics or for detection of error conditions can be implemented in the control. Modeling and diagnosis of various control options as well as error conditions are also involved in our dissertations. We want to simulate these conditions and eliminate their weaknesses before the implementation of the control. The aim is to use the given model in teaching process and realizing various proposals for the control of these manipulators.

References

[1] Jadlovský, J. - Laciňák, S. - Chovaňák, J. - Ilkovič, J.: Proposal of distributed control system of flexible production line (in Slovak), In: Cybernetics and Informatics: International Conference: Vyšná Boča, Slovak Republic, 10th-13th February 2010. - Bratislava: STU, 2010. - ISBN 978-80-227-3241-3. - S. 1-6.
[2] Čopík, Matej: Proposal and implementation of flexible production system control. Diploma thesis. Košice: Technical University of Košice, Faculty of Electrical Engineering and Informatics, 2010. 74 s.
[3] Catalog: Elsys, s.r.o., DeviceNet – Omron basic industrial network, AT&P journal, 8/2004.
[4] Catalog: ILS2D DeviceNet, Lexium Integrated Drive, Product manual, V2.01, 11.2008

Recent Progress in Language Modeling and Continuous Speech Recognition of the Slovak Language

Ján STAŠ, Daniel HLÁDEK

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[jan.stas, daniel.hladek]@tuke.sk

Abstract—Nowadays, speech technologies are very popular and simplify the human-machine interaction as well as the communication between people. In the case of the speech recognition systems, design and complexity of all such systems strongly depends on the given language. Therefore, a creation an efficient language model for a high inflective Slovak language has been a big challenge for us. In this paper we aim to describe recent progress in the modeling the Slovak language for the domain-specific task of the speech recognition. Individual techniques and approaches in the field of text preprocessing, vocabulary selection, language modeling, model adaptation and optimization are described, which stand for the optimal model of the Slovak language. The experiments are focused not only on the evaluation of the language model itself but also on evaluation of our speech recognition system as a whole. We have achieved significant improvement in the domain-oriented modeling of the Slovak language and large vocabulary continuous speech recognition during the last few years.

Keywords—language modeling, model optimization, speech recognition, vocabulary selection

I. INTRODUCTION

Modeling of the Slovak language, which belongs among highly inflective languages, is more difficult than a creation of a model of the English language. First reason is that the Slovak language is characterized by a relative free order of words in sentences that consequently lead to the problem of *data sparseness* of the text data used for training language models (LMs). Second reason is the inflection in the language itself that leads to a several times larger vocabulary than in English. Therefore, it is necessary to collect such amount of text data that could statistically enough cover the given language.

Contemporary modeling of the Slovak language is based on the knowledge of modeling of the related Slavic languages [1], such as Czech, Polish, Serbo-Croatian or Russian language. From the field of statistics, Slovak language is very similar to the Czech language, especially in forming words into sentences and determining the sentence semantics. In the contrast, from linguistic point of view, mainly in phenomena of inflection and assimilation in voice, Slovak is more or less similar to the Polish. Therefore, it is appropriate for the statistical language modeling to be limited with linguistic constraints as well.

This article describes a brief overview of the modeling the Slovak language for domain-specific task of judicature in *large vocabulary continuous speech recognition* (LVCSR). During the last years, we have encountered several problems with text preprocessing, selection of the basic statistical methods used in the modeling of the other languages and adaptation into the area of application. Another important part in the process of language modeling has been optimization of the resultant model, which introduced phonetic and linguistic relations between words. These optimization steps have caused an improvement in quality of the Slovak language model as well as recognition accuracy of the LVCSR system itself.

This article is organized as follows. In the Section II and Section III a short overview about text gathering and preprocessing and creation of vocabulary that have been used in modeling of the Slovak language is described. Then the process of language modeling, selection of appropriate smoothing and pruning technique, method for the domain adaptation and some optimization approaches are mentioned in Section IV and Section V, respectively. In Section VI the setup of the Slovak LVCSR system used in testing process is presented. At the end of this article the selected experimental results for domain-specified task of Slovak LVCSR in Section VII are summarized. Section VIII concludes this article.

II. TEXT CORPORA

The main assumption in the process of creation an effective LM for any inflective language is to collect and consistently process a large amount of text data that enter into the process of training LM. Therefore we have proposed an automatic system, called *webAgent* [2], which retrieves the text data from various web pages and electronic documents that are written in Slovak language. The text data have been preprocessed by the additional modifications such as word tokenization, sentence segmentation, abbreviations expanding, numeral transcription, etc. This system also includes filtering grammatically incorrect words, duplicity verification of text documents and some other constraints. When preprocessing the domain-specific text data we had to resolve the problem with transcription a large amount of specific abbreviations, numerals, dates, and symbols too [3]. Nowadays, we are working with the size of about 2 billion of tokens contained in text corpora divided into several different domains. It should be noted that for filtering of grammatically incorrect words we have used our spellcheck lexicon which was created by merging available Open Source dictionaries such as *aspell*, *hunspell*, and *ispell* with lists of proper nouns and various name entities available on the Internet. The size of our lexicon for spell-checking is about 1.25 million of unique words [4].

TABLE I
VOCABULARY.

Label	Note	Words
245k	general domain (AF select.)	245 310
248k	+ anotations with noise events	248 591
312k	+ geographic name entities	
	+ general domain (ML select.)	312 600
330k	+ judicial domain (ML select.)	330 016
348k	+ general domain (ML select.)	348 480
names	female (inflect. forms)	1 060
	male (inflect. forms)	824
surnames	female (inflect. forms)	55 774
	male (inflect. forms)	82 388
name	geographical	22 050
entities	institutions	2 331
mwe	multi-word expressions	3 182

III. VOCABULARY

Vocabulary that have been used in language modeling was selected using standard method based on *highest occurrence* words in text corpus and *maximum likelihood* approach for selection domain-specific words [5]. The vocabulary was then extended to geographical names and names of various institutions in Slovak republic. We have also proposed an automatic tool for generating inflective word forms for names and surnames which were later used in creating a language model of names and surnames for Slovak LVCSR system. The progress in the increasing the size of vocabulary can be seen in the Table I. We found that the optimal results in the modeling of the Slovak language are achieved when the size of vocabulary is about 100–150 thousand of words for the domain-specific task and 300–350 thousand of words in general domain of speech recognition. It should be noted that all words in vocabulary were manually checked and corrected by linguistic experts.

IV. LANGUAGE MODELING

Statistical modeling of the Slovak language is based on the standard methods for smoothing, pruning and domain-oriented model adaptation as have been used in similar languages. In following sections, these selected methods will be briefly mentioned.

A. Smoothing

As it was mentioned earlier, for dealing with problem of data sparseness, smoothing techniques are usually used in the statistical language modeling. The problem of zero probabilities that lead to the errors in the recognition, smoothing process resolved by uniform redistribution of parts of probabilities of observed n–grams among n–grams that are not observed in training set. Nowadays there exist several different smoothing techniques which use n–grams counts or counting this counts in computation of discount constants. Among all smoothing techniques *the modified Kneser-Ney algorithm* [6] produces optimal results in the modeling of the Slovak language.

B. Pruning

Another problem in the language modeling for the task of real-time application is the size of the resultant LM. With using a large vocabulary increases the number of n-grams in LM that may occur in the training set just once or twice and do not have a big impact on the quality of LM. Therefore, these n–grams can be excluded from the LM using pruning. There exists several criteria for pruning LM. However, the best result was achieved using *pruning based on relative entropy* for any language [7].

C. Adaptation

In the process of enhancing the performance of the ASR systems, the *language model adaptation* (LMA) plays an important role, especially in case of domain-specific speech recognition. The basic idea of the LMA is to use a small amount of domain-specific text data to adjust the LM and reduce the impact of language differences between the training and testing data and set the parameters for LM to correspond as much as possible with the real conditions of LVCSR application. LMA includes not only statistical dependencies between words in given language, but also a frequency of word occurrences, structure of the text and further additional information that usually come from the field of linguistics and phonology. We have observed that the methods which produce significant results for strong statistically dependent languages such as English, do not bring notable improvement in modeling Slovak language. Therefore, usage of the *linear interpolation* or its generalized alternative [8] for Slovak is more than sufficient and interpolation weights should be adjusted using *expectation-maximization* (EM) algorithm [9].

V. MODEL OPTIMIZATION

Several different techniques and approaches have been used and proposed in order to get an efficient model of Slovak language for off-line and on-line speech recognition. Some important ones are described in this section.

A. Annotations and Noise Events

Annotations of the acoustic data used in training acoustic model are a special part of the text corpus. These data also contain a large amount of tags of noise events together with useful text. We have discovered that noise events have a positive effect on the quality of LM, because a speaker often uses noises such as breath, cough, munching, waver, etc. in spontaneous speech. Therefore, we have decided to include these annotations with noise tags into training process and dictionary. It is important to say that after recognition by LVCSR system noise events have appeared like transparent words [3].

B. Spelling Pronunciation

Acronyms, abbreviations, name of titles or single letters are usually spelled letter-by-letter. For a better distinction and separation we have mapped about 600 such events to uppercase and reduced number of pronunciation variants in the dictionary [3].

C. Model of Names and Surnames

In the on-line dictation LVCSR system, most common errors occur in recognition of name entities, especially in case of names, surnames and addresses. Therefore we have created independent model of names and surnames that can be used in the special dictation mode in our LVCSR system. With an increasing number of inflected forms for these name entities, we will consider to use *class-based models* for this purpose, which better model the surrounding context then independent parallel model.

D. Model with Multi-Word Expressions

Besides errors in the recognition name entities, some errors can be seen in recognition of short words at the beginning of the speech or after long pause. To avoid this problem, it is suitable to model these events using *multi-word expressions* [10]. For the process of selection, these multi-words we have used a standard statistical measures based on absolute and relative co-occurrence of these word pairs in the text corpora limited by the linguistic constraints. Experiments for Slovak language has showed that the multi-word expressions do not improve the overall precision of the LVCSR system, but an increase in recognition accuracy that has been observed just in recognition of short words. Usage these models in real-time application can be submitted by the appropriate selection of these units from specific domain.

E. Morphology

The inflection in Slovak language usually occurs on the border of the stem and ending. This knowledge can help in modeling of unknown words or words with a low occurrence in training corpus using *morpheme-based models*. Dividing singletons or words with a low frequency in the training corpus into morphemes, it is statistically possible to cover such events that do not occur in LM. The knowledge of morphology of the given language allows us to also generate new word forms, for example as it was in the case of declination of names and surnames described above.

F. Augmentation Statistics of n–grams

Nowadays, research in the language modeling is oriented on the augmentation of statistics of bigrams and trigrams from other resources than by gathering a large amount of text data of given language. Statistics of seen or unseen n–grams can be obtained either using web search engines [11] or by translation n–gram statistics from other (similar) languages, etc. However, obtained occurrences have to be renormalized regarding occurrence of individual n–grams in training corpus.

VI. LVCSR Setup

The experimental results were performed with trigram LMs created using *The SRI Language Modeling Toolkit* [12] with vocabularies of different size and smoothed by using the *Kneser-Ney algorithm*. For combination and adaptation LMs, standard *linear interpolation* have been used, where interpolation weights were adjusted to the domain of judicature using *expectation-maximization* algorithm. Finally, the resulting LMs have been pruned using an algorithm based on *relative entropy* in order to use them also in real-time application of Slovak LVCSR. Note that LMs used in all experiments had approximately equal number of bigrams and trigrams.

Also the triphone context-dependent acoustic models based on *hidden Markov models* (HMM) have been used, where each state have been modeled by 32 Gaussian mixtures. The model have been generated from feature vectors that contains 39 *mel-frequency cesptral* (MFC) *coefficients*. It has been trained using about 50, 70 and 70 hours of anotated speech recordings of the parliament speech and dictated or imitated judicial proceedings, respectively. The training set involves model of silence, short pause and addtional noise events [13].

For the speech recognition, *high-performance LVCSR engine* [14] with recognition algorithm based on the two-pass strategy has been used. In this case, the input data are processed in the first pass with bigram LM, and the final search for trigram model is performed again using the result of the first pass to narrow the search space [14].

The test data for all experiments were represented by the recordings of judicial proceedings and contain 41 878 words in 3 426 sentences and phrases. We have decided to also use phrases in our test set, because in real conditions, people do not only break on a sentence boundaries, but also on phrase boundaries, usually before conjuction.

For evaluation of our LVCSR system the *word error rate* (WER) evaluated on the test set have been used. WER is a standard measure of the preformance of the ASR system, computed by comparing reference text read by a speaker against the recognized result and takes into account insertion, deletion and substitution errors.

VII. Experiments

Any LVCSR system is a complex system that consists of several parts and each of these parts contributes to the overall recognition result. Similarly it was in our case, when the result of testing the LVCSR system varied depending on the conditions of training acoustic and language models at the same time. With increase of database of the text data used for training language models, the number of hours of annotated recordings of acoustic data also has been increased. Also, an increasing number of verbal forms in dictionary for speech recognition has been conditioned by the amount or availability of training data from domain-specific as well as from general area of speech recognition. In term of time and computational complexity of the training and testing process it was not possible to monitor the contribution of any change in the size of training databases and additional adjustments in training process to the recognition accuracy of our LVCSR system. Therefore, the result in WER of our LVCSR system depicted in the Table II is the result of contributions from individual models, size of vocabulary and language model optimization methods at the same time.

Experiments were performed using only one test set, which statistics were described in previous section. Also, as it was mentioned in the Table I, the size of the vocabulary increases with size of text data used in training LMs. For each word in a vocabulary a phonetic transcription based on SAMPA phonetic alphabet with relevant alternative pronunciations were generated [15]. It should be noted that all words before adding into the vocabulary used in the training LMs or dictionary with a phonetic transcription used in the LVCSR system have been checked by linguistics experts.

TABLE II
EXPERIMENTAL RESULTS IN THE OFF-LINE TESTING OF THE SLOVAK LVCSR SYSTEM FOR THE DOMAIN-SPECIFIC TASK OF JUDICATURE.

AM train [record.]	LM train [tokens]	Vocab [words]	Note	WER [%]
50h	0.5 $billion$	245k		16.70
120h	0.5 $billion$	245k		15.35
120h	0.5 $billion$	248k	+ LM from annotations with noise events	14.50
140h	1.2 $billion$	312k	+ LM from anot. with noise events	11.45
140h	1.6 $billion$	330k	+ LM from annotations with noise events	
			+ LM adaptation	7.45
190h	1.8 $billion$	348k	+ LM from annotations with noise events	
			+ LM adaptation	5.15

In the first two lines in the Table II we can see the influence of the extension of the acoustic database. With more than twice as many of training data and constant parameters of the language model and vocabulary we have achieved almost 1.5 % decreasing in WER of the LVCSR system. By combining a general LM with LM created from the text data of annotated speech recordings used in training acoustic model (AM), we have improved the recognition accuracy by 0.85 %. Subsequent extension of the acoustic and text data in the training AM and LM and adding new words into the vocabulary have increased the recognition accuracy by more than 3 %. Further addition of some new words and text data from the field of judicature and strong LM adaptation to this domain has achieved significant improvement in recognition accuracy of the LVCSR system. Repeated enlargement of the acoustic and text data, adding some specific words from the perifery of the vocabulary and re-training AM and LM we reached the 5.15 % WER and approximately 0.5 % *out-of-vocabulary* (OOV) words evaluated on test data set. Further improvement could be observed by using a *gender-dependent* or *gender-independent acoustic model adaptation* in Slovak LVCSR system which is still in a researching stage [16].

Finally, it should be noted that our dictation LVCSR system works successfully also in the on-line mode. The most common errors using this LVCSR system occur in recognition name entities, due to the large amount of these unique base as well as their inflected word forms. However these errors do not have a big impact on the overall misrecognition of the LVCSR system.

Regarding to the current recognition accuracy of the LVCSR system, the field of modeling of the Slovak language we would like to focus on potencial usage of *class-based models* for better modeling of the name entities. Our research also should be focused on the reducing errors in the recognition of the short words at the beginning of the speech or after a long pause using *multi-word expressions* and discover an influence of the *morpheme-based* models in the modeling of unseen or unknown words in Slovak LVCSR. This knowledge will be subject for further research.

VIII. CONCLUSION

In this article a brief summary of methods and approaches used in modeling of the Slovak language have been presented. Combination of standard methods of statistical language modeling and optimization techniques, that bring into the training process of LM an additional information, often linguistic regularities as well, we achieved notable improvement in recognition accuracy of LVCSR system. In future work, we want to focus on morphologically motivated class-based models and morpheme-based models to better model specific cases such as name entities and addition of unknown words into the model in on-line application of Slovak LVCSR system.

ACKNOWLEDGMENT

REFERENCES

[1] J. Nouza, J. Zdansky, P. Cerva, and J. Silovsky, "Challenges in Speech Processing of Slavic Languages (Case Studies in Speech Recognition of Czech and Slovak)," *A. Esposito et al. (Eds.): Development of Multimodal Interface: Active Learning and Synchrony, Springer-Verlag, LNCS 5967*, pp. 225–241, 2010.

[2] D. Hládek and J. Staš, "Text Mining and Processing for Corpora Creation in Slovak Language," *Journal of Computer Science and Control Systems*, vol. 3, no. 1, pp. 65–68, 2010.

[3] J. Staš, M. Trnka, D. Hládek, and J. Juhár, "Test Preprocessing and Language Modeling for Domain-Specific Task of Slovak LVCSR," in *Proc. of the 7th Intl. Workshop on Digital Technologies, Circuits, System and Signal Processing (DT 2010)*, Žilina, Slovakia, 2010, pp. 1–4.

[4] J. Staš, D. Hládek, M. Pleva, and J. Juhár, "Slovak Language Model from Internet Text Data," *A. Esposito et al. (Eds.): Towards Autonomous, Adaptive, and Context-Aware Multimodal Interfaces: Theoretical and Practical Issues, Springer-Verlag, LNCS 6456*, pp. 352–358, 2010.

[5] A. Venkataraman and W. Wang, "Techniques for Effective Vocabulary Selection," in *Proc. of EUROSPEECH*, Geneva, 2003, pp. 245–248.

[6] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd Edition)*. New Jersey: Prentice Hall, Pearson Education, 2009.

[7] A. Stolcke, "Entropy-based Pruning of Backoff Language Models," in *Proc. of DARPA Broadcast News Transcription and Understanding Workshop*, Lansdowne, 2002, pp. 270–274.

[8] B. J. Hsu, "Generalized Linear Interpolation of Language Models," in *IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2007)*, Kyoto, Japan, 2007, pp. 136–140.

[9] J. Staš, D. Hládek, and J. Juhár, "Language Model Adaptation for Slovak LVCSR," in *Proc. of Intl. Conf. on Applied Electrical Engineering and Informatics (AEI 2010)*, Venice, Italy, 2010, pp. 101–106.

[10] J. Kolorenč, J. Nouza, and P. Červa, "Multi-words in the Czech TV/radio News Transcription System," in *Intl. Conf. Speech and Computer (SPECOM 2006)*, St. Petersburg, Russia, 2006, pp. 70–74.

[11] F. Keller and M. Lapata, "Using the Web to Obtain Frequencies for Unseen Bigrams," *Computational Linguistics*, vol. 29, no. 3, pp. 459–484, 2003.

[12] A. Stolcke, "SRILM – An Extensible Language Modeling Toolkit," in *Proc. of the 7th Intl. Conf. on Spoken Language Processing (ICSLP 2002)*, Denver, Colorado, 2002, pp. 901–904.

[13] R. Sabo and S. Darjaa, "Automatický prepis rokovaní Národnej rady Slovenskej republiky – rečová databáza, akustické modely, rozpoznávanie reči," in *Proc. of the 15th Intl. Acoustic Conference Noise and Vibration in Practice*, Kočovce, Slovakia, 2010, pp. 107–110.

[14] T. Lee, T. Kawahara, and K. Shikano, "Julius – An Open Source Real-time Large Vocabulary Recognition Engine," in *Proc. of EUROSPEECH*, 2001, pp. 1691–1694.

[15] M. Rusko, M. Trnka, and S. Darjaa, "MobilDat-SK – A Mobile Telephone Extension to the SpeechDat-E SK Telephone Speech Database in Slovak," in *Proc. of the 11th Intl. Conf. Speech and Computer (SPEECOM 2006)*, Sankt Peterburg, Russia, 2006.

[16] M. Papco and J. Juhár, "MLLR Adaptation of Speaker Independent Acoustic Model to a New Speaker in LVCSR," in *Proc. of the 7th Intl. Workshop on Digital Technologies, Circuits, System and Signal Processing (DT 2010)*, Žilina, Slovakia, 2010, pp. 1–4.

Round Trip Time distance estimation for WLAN 802.11 devices

Vladimír CIPOV

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

vladimir.cipov@tuke.sk

Abstract—**In most cases the precise distance estimation between the wireless nodes is one of the most important steps for successful localization. Especially in urban and rural environment with numerous reflectors and obstacles it represents a considerable problem. In this paper the Round Trip Time distance estimation technique is presented. It is implemented into the trilateration-based positioning algorithm introduced before with using RSS and mutually compared. The purpose of this comparative study was to show how the location accuracy will be increased using the time of flight technique instead of received signal strength. This paper is structured as follows. At the beginning two main distance estimation opportunities are briefly described. Next the Round Trip Time estimation techniques are introduced. The simplified RTT model is also presented and implemented into the localization algorithm. At the end the simulation results of this comparative study are concluded.**

Keywords—**Round Trip Time, WLAN, ToF, ranging.**

I. INTRODUCTION

An essential part of the range-based [5] localization algorithms is to estimate the distances between the pairs of wireless nodes. In general a few techniques have been considered. Two most used are called as Received Signal Strength (RSS) and Time of Flight (ToF) [4]. The basic signal strength measurement and its conversion to the distance is not a reliable localization parameter especially for hard urban or indoor environment with numerous obstacles. Measurements are impacted by both time-varying errors and environmental-dependent factors. In most cases these factors are unpredictable and are considered as a random variable [4]. Due to these drawbacks and due to the fact that signal strength sharply decreases in a non-linear and unpredicted fashion with distance, increasingly ToF distance estimation technique becomes an object of the research activity. Two approaches of the ToF measurement can be used.

- the classic approach which measures the *time of pure radio signals* and
- *Round Trip Time* (RTT) measurement using standard WLAN packets.

In the first case applying special signal processing algorithms based on cross-correlation techniques must be considered and the time synchronization of WLAN devices should be assigned [6]. Also difficult but more promising approach for estimating the distance between two nodes is represented by RTT measurement which uses the standard WLAN packets and is in detail described in the following section [4].

II. ROUND TRIP TIME DISTANCE ESTIMATION PRINCIPLE

In general, the distance between two nodes (*d*) can be obtained by multiplying the signal propagation time (*t*) by the speed of light (*c*=3.10^{8}m.s^{-1}):

$$d = c.t \qquad (1)$$

In case of the Round Trip Time distance estimation techniques *t* can be also expressed as RTT parameter which represents the time the signal spends travelling from a transmitter to a receiver and back again [1]. Then the pure time of flight of the propagated signal is obtained by halving the whole propagation time RTT and this fact can be expressed as [1]:

$$d = c.\left(\frac{RTT}{2} \right) \qquad (2)$$

The described fact represents the simplest approach to the time of flight estimation using Round Trip Time method.

In fact the data exchange and the following time estimation can be explained using standard 802.11 with RTS/CTS (Request-to-Send/Clear-to-Send) mode. Link-layer data and ACK frames can be used. Then the RTT estimation represents the time elapsed between sending a link-layer data frame to the reference device and a consequent reception of the link-layer ACK frames in the mobile device. These are a frame sent by the transmitter and an answering frame from the receiver [1][2]. Now it can be written that RTT duration consists of the pure time of signal *propagation* between the transmitter and receiver (and vice versa) and the *processing* time at the receiver side [1]:

$$RTT = t_{prop\_TR} + t_{proc} + t_{prop\_RT} \qquad (3)$$

where t_{prop_TR} represents the propagation time of the data frame from the transmitter to the receiver, t_{proc} the processing time of the data frame at the receiver side and t_{prop_RT} the propagation time of the ACK frame from the receiver back to the transmitter. It is assumed that time duration of the t_{prop_TR} and t_{prop_RT} will be the same. Due to this assumption the relationship can be rewritten as [1]:

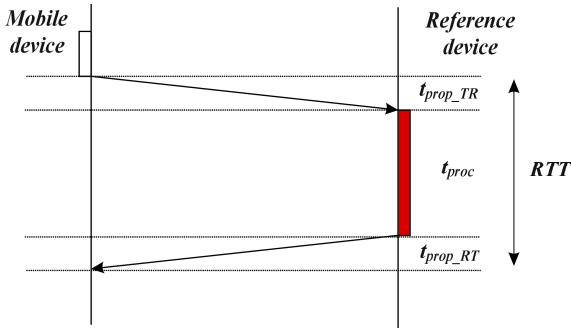$$RTT = 2.t_{prop} + t_{proc} \qquad (4)$$

Fig. 1. Basic Round Trip Time principle

To calculate the distance between transmitter and receiver t_{proc} should be isolated from the whole RTT. f_{clock} represents the frequency of the receiver clock [1]:

$$d = c \cdot \left(\frac{RTT - t_{proc}}{2} \right) \cdot \left(\frac{1}{f_{clock}} \right) \qquad (5)$$

Based on the approach described above the Two-way measurement principle can be carried out [3]. As was mentioned, if considering IEEE 802.11 standard the Two-way measurement principle can be applied using the packet pairs DATA and ACK. If the transmission duration of the data packet, the time stamps at the start of transmission of the data packet and at the start of receiving the acknowledgment packet are known the pure time-of-flight can be expressed as [3]:

$$t_{tof} = \frac{1}{2}\left(t_R - t_T - t_{DATA} - t_{SIFS} \right) \qquad (6)$$

where t_R and t_T represent the time stamps at the transmission and reception of the DATA packets, t_{DATA} represents its time duration and t_{SIFS} represent the time duration of the SHORT INTERFRAME SEQUENCE (10µs/16µs).
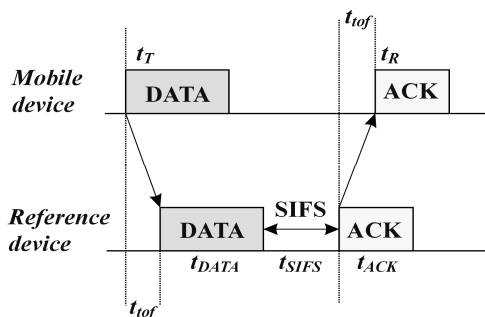
Fig. 2 Two-way Round Trip Time measurement principle using packet pair DATA/ACK

In [3] also the Four-way measurement principle is initiated. The measurement principle can be extended by using multiple packets sent in a packet sequence of which transmission is supported. After the RTS, CTS, DATA and ACK packets are selected then the time-of-flight can be expressed as [3]:

$$t_{tof} = \frac{1}{4}\left(t_{3\_MD} - t_{0\_MD} - t_{DATA} - t_{CTS} - t_{RTS} - 3t_{SIFS} \right) \qquad (7)$$

where the parameters of the equation (7) are the similar as in

case of the Two-way procedure and they are visible on Fig. 3.
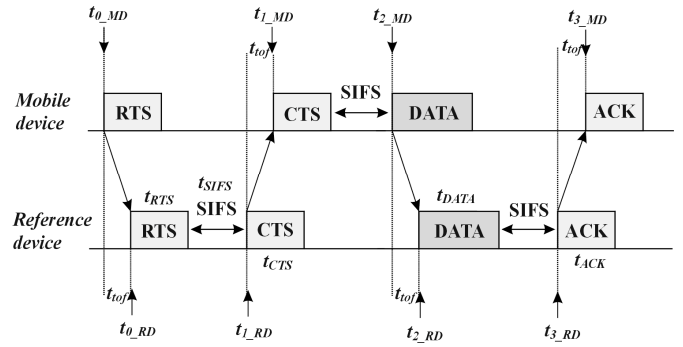
Fig. 3 Four-way Round Trip Time measurement principle using multiple packets sequence

RTT is highly time-variant due to the impact of multi-path signal propagation and clock quantification errors [2]. In terms of accuracy, RTT measurements should ideally be performed at the physical level [3]. In [1] authors use an auxiliary hardware subsystem connected to the WLAN card and to the parallel port in the mobile device which provides clock counting using the built-in 44 MHz clock. Such hardware is needed because of the overall RTT is expected to be in the order of microseconds which is not achievable using standard software.

III. TIME ESTIMATION MODEL

The model for determination the time of flight of the propagated signal can be seen on Fig. 4.
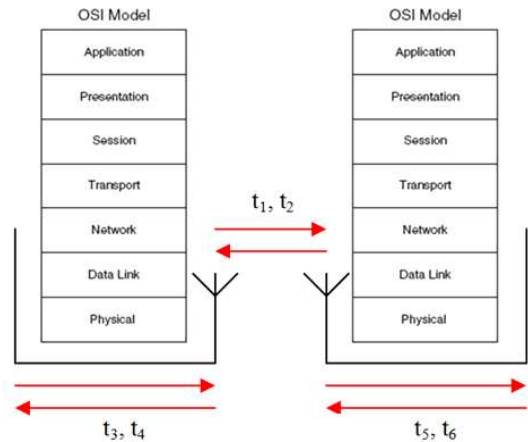
Fig. 4 Time estimation model

As is shown the whole Round Trip Time consists of several time components. Two sets which cause the time delay of the received signal and influence the accuracy of the distance estimation were considered [1].

• First, which influence the time components t_1 and t_2, related to signal propagation represented by:
   o   multi-path propagation and
   o   speed of light which can be a little bit different instead of the speed of light in vacuum (3.10^8m/s).

• Second, which influence the rest time components, related to signal processing represented by:
   o   discrete time quantification,
   o   delay due to the hardware electronics,
   o   drift of the card clock in the mobile device during

- measurement,
- o relative drift between the clocks in the mobile and reference device,
- o short interframe sequences.

IV. SIMULATIONS AND RESULTS

The simplified time estimation model was implemented into the trilateration-based localization algorithm which was introduced in [7] using RSS measurement. These measurements were simulated for both the rural and also urban environment. The reached results can be seen in TABLE II.

TABLE I
DISTANCE ESTIMATION ERROR USING DIFFERENT Δt_2

$\Delta t_2 [s]$	MIN[m]	MAX[m]	$\emptyset$[m]
10^{-12}	$0{,}4{.}10^{-3}$	6,814	3,450
10^{-11}	0,0066	6,816	3,535
10^{-10}	0,0328	6,854	3,402
10^{-9}	0,1122	7,752	3,502
10^{-8}	0,2623	17,055	5,917
10^{-7}	0,6663	101,56	27,12
10^{-6}	1,3834	1070,7	252,9
10^{-5}	9,0304	12645	2431,2

TABLE II
COMPARISON OF LOCATION ACCURACY USING BOTH RSS AND RTT
MEASUREMENT

LOCATION ACCURACY USING **RSS** MEASUREMENT	
	Location accuracy [m]
RURAL	100 – 280
URBAN	290 – more (unusable)

LOCATION ACCURACY USING **RTT** MEASUREMENT	
Time delay of Δt_2 [s]	**Location accuracy [m]**
10^{-12}	2,15 (0,72%)
10^{-11}	1,98 (0,66%)
10^{-10}	2,14 (0,72%)
10^{-9}	2,36 (0,79%)
10^{-8}	3,98 (1,33%)
10^{-7}	12,23 (4,11%)
10^{-6}	53,93 (17,98%)
10^{-5}	330,68 (110,2%)

For using RTT all of mentioned time components were substituted using two time delay parameters, $\Delta t = \Delta t_1 + \Delta t_2$, where Δt_1 represented the time delay caused by receiver clock counter. For this parameter 44MHz clocks was simulated which caused the fundamental estimation error (0m-6,82m). These values were generated using uniform probability distribution in interval $<0s-(1/44.10^6s)>$. The parameter Δt_2 represented the time delay caused by multi-path propagation and was generated as absolute value from Gauss probability distribution with mean μ=0 and dispersion σ=10^{-12}s, 10^{-11}s, 10^{-10}s, 10^{-9}s, 10^{-8}s, 10^{-7}s, 10^{-6}s, and 10^{-5}s. These values represented the probability of the LoS (*Line-of-Sight*) or NLoS (*Non Line-of-Sight*) communication. The reached location accuracy and its comparison with the location accuracy using RSS can be founded in TABLE II. The results expressed using "%" was expressed according to the terminals radio range (300m). All of the network parameters such as network area, number of unknown and reference nodes can be founded in [7] and for the relevance of the comparative study are the same. For statistical relevance all of the simulations were hundred times repeated.

TABLE I shows the values of the distance estimation errors which was reached using different Δt_2 parameter. These values were thousand times generated and the minimum, maximum and average values were noticed.

V. CONCLUSION

Due to the limitations of current hardware and protocol configuration, there are the problems with measuring and especially with synchronizing time for WLAN [4]. Thus the ToF does not currently work well for WLAN localization systems. However some researchers in their publications investigate the new opportunities of the ToF implementation using the existing WLAN structures or using an auxiliary hardware. These authors reach also good results of the distance estimation, often about from 1m to 4m. The results presented in this paper show the increasing location accuracy and promise the hopeful future implementation of the presented positioning algorithm especially for urban or indoor localization using RTT.

REFERENCES

[1] M. Ciurana, F. Barceló, F. Izquierdo, "A ranging system with IEEE 802.11 data frames", IEEE RWS 2007, Long Beach (USA), January 2007.

[2] F. Izquierdo, M. Ciurana, F. Barcelo, J. Paradells, E. Zola, "Performance evaluation of a TOA-based trilateration method to locate terminals in wlan", in 1st International Symposium on Wireless Pervasive Computing, 2006, p. 1.

[3] Ch. Hoene, J. Willmann, "Four-way TOA and software-based trilateration of IEEE 802.11 devices", in IEEE PIMRC, Cannes, September 2008.

[4] K. Muthukrishnan, G.T. Koprinkov, N. Meratnia, M.E.N Lijding, "Using time-of-flight for WLAN localization: feasibility study", Technical Report TR-CTIT-06-28, Centre for Telematics and Information Technology, University of Twente, Enschede. 2006. ISSN 1381-3625.

[5] A. Srinivasan, J. Wu, "A Survey on Secure Localization in Wireless Sensor Networks", Encyclopedia of Wireless and Mobile Communications, B. Furht (ed.), CRC Press, Taylor and Francis Group, 2008.

[6] A. Günther, Ch. Hoene, "Measuring round trip times to determine the distance between WLAN nodes", in Networking 2005, Waterloo, Canada, May 2005.

[7] Ľ. Doboš, V. Cipov, "Beacon-Based Localization Algorithm for MANET Using Received Signal Strength Measurement", in Proceedings of IEEE International Conference MCSS 2010, Krakow, Poland: 6.-7. May 2010. ISBN 978-83-88309-92-2, pp. 35-40.

Semantic Network – Synonyms and misspelling support

[1]*Stanislav Dvorščák,* [2]*Kristína Machová*

[12]Department of Cybernetics and Artificial Intelligence, FEI - Faculty of Electrical Engineering and Informatics, TU-Technical University of Košice, Letná 9, 042 00, Košice

[1]dvorscaks@openportal.sk, [2]kristina.machova@tuke.sk

Abstract—**Articles introduces one property of implemented semantic network. It is about failure tolerance (misspelling) and synonyms inside documents.**

Keywords—**misspelling tolerance, synonyms, semantic network**

I. INTRODUCTION

A. Project introduction

We are working on the intelligent search machine with the Semantic Support [3]. It is implementation of J2EE application which is composed from multiple modules: harvester, document preprocessing, semantic network, inference mechanism, search module.

The search mechanism it is based on the semantic search which is alternative to the keywords search. It works with semantic information on the documents side. Also contextual information and a lot functionality it is realized in the fuzzy form, which provides the ability to be tolerance to different forms of the same words (synonyms), but also to be tolerant to different spelling mistakes.

B. Article introduction

The input of the application it is based from input data provided by the people. It means these data contain different spelling mistakes and synonyms. What it is typical for input provided by the people, but the same is also for data which are received from the machine measurements [4].

These input (mistakes) can be simple ignored. But on other side these information is also information, everything what it is necessary to do, it is just to find the correct meaning of the information. And sometimes these information are more important as any other correct information.

We are providing these freedom for users of our system. And we want to have tolerant system to using of synonyms and also to be tolerant to different spelling mistakes.

This article provided the clarification of our solution via semantic network. Which is tolerant to mentioned spelling mistakes and synonyms.

II. SEMANTIC NETWORK

A. Standard "Semantic network"

The semantic network is often the knowledge representation. It can be in the form of directed or undirected graph. In the common it is realized as the directed graph with the relations (edges) between the objects (vertices).

These common form can be defined in different forms e.g.: triplets.

B. Our implementation

In our case the semantic network can be imagine as the graph of symbols (vertices) with relations (edges) between them. The symbol is elementary thing of the information stored in the our semantic network. It can be character, word, sentence, some abstract thing, etc... On other side we have the edges which represent the relations between the symbols.
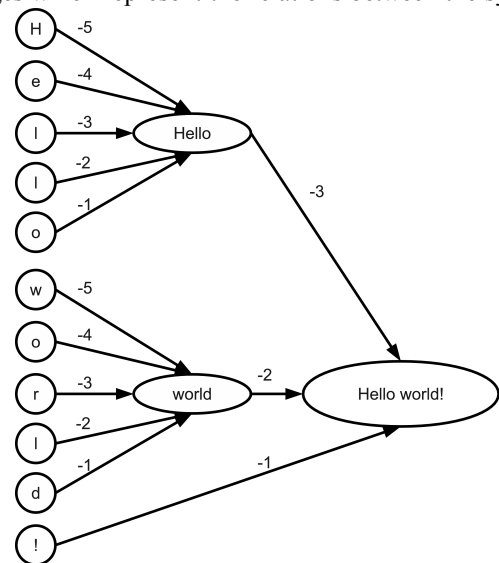
Fig. 1.: Semantic network for "Hello world!"

On the Fig.1. can be founded the representation of semantic network for phrase "Hello world!". It is composed from characters. These character symbols it is connected to the two words symbols and to the one sentence symbol.

From common semantic network representation we have multiple extensions.

A formal semantic network formula is: N = (S, C), where S is the set of symbols, objects (nodes) and the C is the set of connections (edges). Because it is possible to have multiple connections to the same symbol, the aggregation function have to be defined – f_a.

Connection (c) is defined:

S_{input} .... input symbol

S_{output} .... output symbol

$t_{delay} = 1,...,n$ ...... time delay of signal transition, where $t_{delay} > 0$.

e_{inv} .... flag is inverted or is not inverted.

Symbol (s) is defined:

f_a …..................... aggregation function OR/AND

x …..................... strength of signal $<-1, 1>$.

The flag inversion represents negation of the signal. The greater (smaller) validity of the input signal induces the smaller (greater) validity of the output signal [3].

III. AGGREGATION

Aggregation function can be realized in the different way: by AND and OR functions. For AND aggregation the minimum function can be used. The minimal input signal is selected for propagation. But the implementation of this approach is very strict. For reasoning, the geometric weighted distance is used for aggregation. This access provides better results. In other words, two vectors are considered. Both of them are located in the coordinate system. The start point of each of them is represented by point [0, 0, ... , 0]. The second point is represented by the signal values of aggregated symbols. The end point of one vector is represented by the real values of signals. Second vector which represents the full validity is represented by the point [1, 1,..., 1].

$$s_{input} = \langle s_1, s_2, ..., s_n \rangle \, true = \langle 1,1,...,1 \rangle \|true\| = n$$

$$s_{out} = |s_{input}| \cos(s_{input}, true)$$

$$\cos(u,v) = \frac{u_1 v_1 + u_2 v_2 + ... + u_n v_n}{\sqrt{u_1^2 + u_2^2 + ... + u_n^2} \sqrt{v_1^2 + v_2^2 + ... + v_n^2}}$$

$$\cos(s, true) = \frac{s_1 + s_2 + ... + s_n}{\sqrt{s_1^2 + s_2^2 + ... + s_n^2} \sqrt{n}}$$

The OR aggregation is realized as a maximum function over all signals [3].

IV. CONCLUSION

A. Misspelling

This is one it is the common problem of data provided by the people. This type of mistakes can be discovered on two places. As the input of documents or as the part of search question.

It can be simple ignored on the question side. Because user it is able to provided correction of question. And it can be also in the form of suggestion for correction. But it is still the same it is about the customization of the question into the different form.

But this principle can not be used on the level of documents. Because there is no input from the user side. Indexation of documents is automatic process with minimal input from the user.

As it was mentioned in the example of the semantic network. We are using the characters as the elementary symbol extracted from documents. These characters are connected to the words and words are connected to the phrases, phrases to sentences, etc... This structure of network provides that each time also the words which are around the specified word will be part of inference. There is used kind of backtracking algorithm with trashholding constants because of performance. But the words: "world" and "word" will be fired with similar strength.

B. Synonyms

The second support of the application is tolerance to the synonyms. Normally the people using the synonyms. If the key words based search is used, it is necessary to extend the original question with the synonyms.

The structure of our network it is also defined in the form of ontology[2]. And these ontology provides the relation between the synonyms. It means if the word which contains also the synonyms it is used, these synonyms will be also part of the inference of the semantic network. It is similar to the extension of question, but it is done via semantic network. There is also some rules, which provides support for contextual information. Which represents better power as the simple synonyms support [1].

REFERENCES

[1] Antoniou, G.: Nonmonotonic Reasoning. April 1997, 299 pp. ISBN-10: 0-262-01157-3.

[2] Antoniu, G., van Harmelen, F.: A Semantic Web Primer. Massachusetts Institute of Technology, USA, 2004, 238 pp., ISBN 0-262-01210-3.

[3] Dvorčšák, S., Machová, K.: Intelligent Search Machine with the Semantic Support. Proc. of the 9th annual conference Znalosti 2010, Jindřichuv Hradec, Czech Republic, February 3-5, 2010, Publisher: VŠE v Praze, Praha, 2010, 199-202, ISBN 978-80-245-1636-3.

[4] Návrat, P., Paralič, J.: Acquiring, Organising and Presenting Information and Knowledge on the Web. Computing and Informatics, Vol. 28, No. 4, 393-398, 2009, ISSN 1335-9150.

Software-based computer emulation using GPGPU

[1]*Peter JAKUBČO,* [2]*Eva DANKOVÁ,*[3]*Milan VRÁBEĽ*

[1,2,3]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]peter.jakubco@tuke.sk, [2]eva.dankova@tuke.sk, [3]milan.vrabel@tuke.sk

Abstract—**The article deals with the problematics of imitating a computer behavior, i.e. processor ISA. A new idea about constructing an emulator of a processor using GPGPU technology is presented. The key aim is to improve the performance of emulation, trying to make distributed nature of an emulator. The emulator consists of two parts that communicate together in distributed manner. There are also some real experimental results of emulating abstract Random Access Machine.**

Keywords—**computer emulation, ISA, GPGPU, OpenCL**

I. Introduction

Computer emulators have long history. Mostly they are concerned with arcades and video-games. More known scientific term seems to be the simulation, therefore we first point at the difference between simulation and emulation. In extreme abstraction, we can state that simulator imitates the microarchitecture of a computer, and if the imitation is accurate enough, the behavior will be preserved as a consequence. An emulator, instead, imitates the behavior in the first case and then the microarchitecture may be imitated as adding more details into emulation [1].

In science, mostly simulators are used. The main domains of interest is hardware development, where many simulators can be used in each level of abstraction. Emulators are more appropriate when the behavior is important as a whole, from the point of view of inputs and outputs, where does not matter what internal organisation structure the system has. Therefore, emulators are implemented according to the specification defined as a set of intuitive algorithms. In the computer emulation, emulators implement Instruction Set Architecture (ISA) of a computer [2]. The ISA is a programmer's point of view of a computer, the most part is about a processor. Instructions are defined as small algorithms that operates with programmer-level computer structure (registers, flags, memory cells, input/output ports, etc.).

The most used techniques for both simulation/emulation of a processor are interpretation and binary translation [3]. Each technique used for emulation has its history in compiler theory. The binary translation is more powerful than interpretation; in our experiment implementing Random Access Machine (RAM) [4] emulator, we achieved about 97 MIPS (interpretation took about 9.2 MIPS) using Intel Core2 Duo, 1.66Ghz and 2GB RAM.

When emulating more complex computers, such as those with parallelism as multi-processor systems, the uniprocessor host computer suffers of performance slowdown several times (N processors emulation). If parallel emulation algorithms were used, the parallel computer architectures would be able to perform the emulation without significant performance loss. However, one assumption is that, in order the emulation could be cheap, standard PCs should be used for the emulation. The idea is therefore to utilize natural parallel system that can be included within standard PCs; the components are current GPUs.

II. GPGPU and emulation?

There exist data-based and task-based parallelism. Modern graphic cards provides all the parallelisms, by several stream processors [5] that are included in GPUs, ranging from 80-1500. Stream processors use data-based parallelism naturally, but they allow to use also task-based parallelism. However, the task-based parallelism is limited; details can be found in [6].

Stream processing is very suitable for computational tasks. According to [7], I/O (interactive) tasks do not profit performance improvement. In practical speaking, within GPUs, the performance improvement is achieved by:

1) Minimalization of inputs and outputs;
2) Try to make each computation long enough, spending there as much time as it is possible.

The main task now will be to determine if the work-flow of a CPU has computational nature. Sequential (uniprocessor) CPU executes instructions in an infinite loop. The loop is never interrupted, besides a computer turn-off. Concurrently, devices perform their own work. Besides, they can communicate with CPU, and, possibly between themselves. Sometimes particular components must wait for data that cannot be computed immediately, but are taken as an input from an input device (e.g. input from keyboard). In such way data between particular components "migrate" and the computation is realized in a component at the end of the data path. If all components are emulated by single host platform (the classic approach) and if all data can be computed without any interaction, the CPU work-flow will have computation nature. This is, ofcourse, not true. However, there exist processes that stay longer in computation, with minimal interaction (e.g. video/image/sound processing, weather forecast, etc.)[1]. Therefore, the question about the CPU computational nature can be concretized to a question: 'What kind of tasks the CPU performs?'. The compute-bound or interactive-bound processes have better performance, if the computer is adapted for particular kind of work, respectively.

Some compute-bound processes, and also interactive-bound processes have their computer architectures representations. GPUs rather do computational jobs, that are well parallelized in the meaning of data-parallelism. Then it can be stated that it is worth to emulate only compute-bound computers with a tendency of good parallelization. Examples of such

---

[1]Note that data processing in the computation is not considered as the interaction when the data is located in operating memory.

CPU representations can be found: data-flow architectures, co-processors, or general-purpose processors about that we actually know they will be compute-bound. Examples of compute-bound and well-parallelizable tasks are: video/image/sound processing, some physical model computations, cryptographic tasks, neural network simulation, etc. Emulation of computers that are primarily intended for executing such tasks, would be usually very slow in non-parallel environment. Therefore, using better non-parallel emulation technique will not help to improve the performance.

Next question is about emulator structure organization. Using single-platform, all target computer components are implemented as software modules or parts (e.g. CPU, memory and I/O devices) that communicate with each other in a sense of procedure calls. Physically, they are all located in the host operating memory, so the physical communication is realized only between host CPU and host operating memory. When we want to utilize GPUs as another power source, it is necessary to physically separate the emulator modules. In that point, distributed nature is created. All compute-bound components should be executed in a GPU and the I/O bound components in the host CPU. The reason is that GPU is not connected to another devices, except the host operating memory. The communication between target components implemented in host GPU and target components implemented in host CPU must go through host CPU.

III. APPROPRIATE STREAMING LANGUAGE

Now it is appropriate to choose the implementation language used for emulator GPU implementation. Currently, three options can be considered: CUDA [8], OpenCL [9] and DirectCompute [10]. For further work, we choosed the OpenCL language. The reason is that it is an open standard and can work with both CPUs, and GPUs. OpenCL is supported by many vendors and there exist free SDK. The standard is still under development, so now it has several restrictions. The most critical one is that within ATI SDK [11], independent kernels cannot be executed concurrently. This restricts the task-based parallelism. However, it is worth to expect an improvement of the situatuon, because in the present time, there exist some GPUs that support few concurrent independent kernels execution (CUDA supports maximum 4 independent kernels that can run concurrently on Fermi cards [12]).

IV. EMULATION TECHNIQUE ON GPU

GPGPU is a technique that enables general-purpose computation on graphics cards. It uses a shader language as the lowest-level language (machine language). However, shading languages are mostly enclosed (their specification is not public) and they differ among GPU models. The access is done directly via drivers, or via higher level streaming languages (e.g. OpenCL, CUDA, DirectCompute; older approaches include OpenGL and DirectX).

The portability problems connected with the enclosed shading languages specifications make impossible of performing other technique than interpretation. Another critical problem is, that kernels[2], within all present-available streaming languages, cannot use function pointers. Therefore, only decode-and-dispatch interpreters is possible to construct (instead of faster threaded code interpretes) [1].

[2]Small programs executed in GPU in parallel

V. PROPOSED EMULATOR STRUCTURE AND COMMUNICATION AMONG EMULATED COMPONENTS

The structure of the distributed emulator is depicted in Fig. 1. The emulator consists of two main parts that communicate in the distributed manner. The idea is that GPU would execute the CPU emulator that would be implemented as a stream kernel. The host part will act like scheduler or dispatcher, that would *catch* and *delegate* events either from the target CPU (physically from host GPU), or from other target (emulated) I/O devices (physically from host CPU).
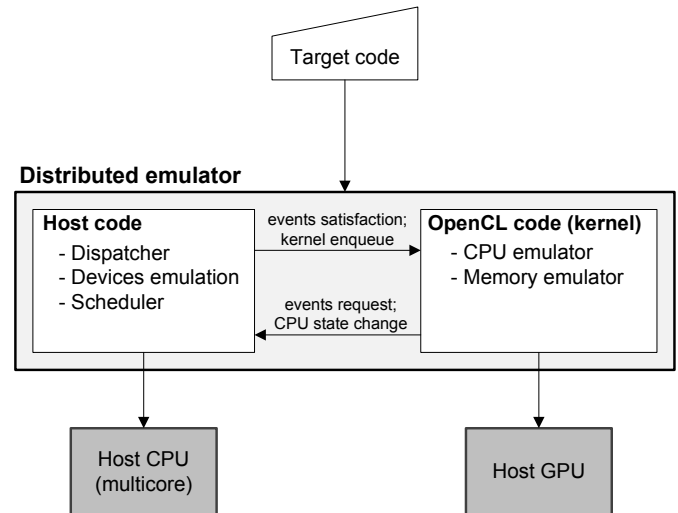
Fig. 1. Structure of distributed computer emulator using GPGPU

As can be seen in the figure, compute-bound target CPU and operating memory are located in host GPU. The I/O devices emulators are considered to be interactive-bound by default, and are located in the host CPU. The communication between host CPU and GPU therefore represents the communication between target CPU and I/O devices.

The physical transmission medium (further a channel) between CPU and GPU is PCI Express channel. Transfer performance is CPU and chipset dependent. Transfers from host CPU to GPU are done either by the command processor or by the DMA engine. The GPU also can read and write host operating memory directly from the GPU through kernel instructions over the PCIe bus.

The emulated CPU, even it would be compute-bound, generally may perform also some interaction. The interaction can be for example a reaction to an external event (e.g. device interrupt request). CPU may also ask for service if it needs to send the output data. All the interaction must go through the host-part of the emulator and therefore, some communication mechanism must be provided.

Some CPUs ask for input manually (using so-called polling, or busy waiting) – in that point of view, CPU emulators generate event requests for inputs and wait for it in loops. Another option is that devices would request for service by so-called interrupt requests (IRQs). The interrupt request in higher abstraction can be understood as an event that is catched by host CPU (using spooling). The CPU reacts to these requests by interruption of current execution and jumping to service procedure. The advantage is that the CPU does not have to wait for the event and can do another work independently. The model of proposed strucutre in Fig. 2 explains the idea using GPGPU.

Fig. 2.   The example communication between host CPU and GPU – events

In the figure, a device asks for a service, by generating an interrupt request. The interrupt request is, however, virtual, because it goes from emulated device. The host-part of the emulator catches the event and processes it. If the interrupt needs to be serviced by the emulated CPU, the interrupt request, along with additional i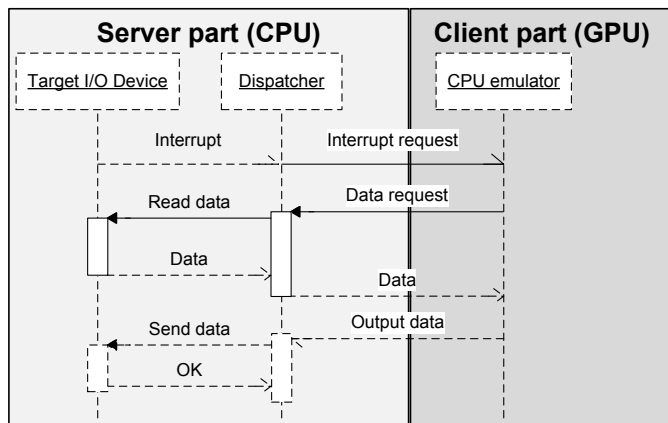nformation, is 'sent' to the GPU. However, it is impossible to interrupt running kernel on the GPU (i.e. using any of of the present streaming languages it is impossible). The solution is therefore to use some other communication mechanism. The OpenCL allows to use shared memory between the host and GPU, located in the host operating memory. As was mentioned earlier, the PCIe bus connects a GPU also with the system memory and the OpenCL API allows to do that programatically. So, the 'send' operation from host CPU into GPU means to change specific data in shared memory by the host-part of the emulator (the 'send' operation in the another direction would be performed in the same manner, but the GPU will perform change of the shared data). Therefore, both GPU kernel and host-part of the emulator must watch the shared memory for changes in an infinite loop. When GPU kernel encounters the change, the CPU emulator immediately interrupts its execution flow and handles the interrupt in specified way. If the host-part encounters the change, according to the negotiated meaning of the data it performs specified action (e.g. send data to an output device).

In such way data request from the emulated CPU is generated. The host-part of the emulator then ensures the gathering of required data for the emulated CPU satisfaction. The other component collects the data and sends it to the host-part of the emulator. The communication mechanism between dispatcher in the host-part of the emulator and other emulated components is not restricted to the shared memory mechanism. The reason is that it is assumed that both the dispatcher and other emulated components are located in host operating memory and executed by host CPU. The host part finally sends the requested data back into the CPU using the shared memory.

The last part of the diagram is showing the other situation – CPU wants to send output to an output device. In the figure, the 'other component' represents both the input and output device. The output device should confirm that the sent data were accepted correctly.

VI. Implementation notes about CPU emulator

Main issues of the implementation of distributed emulator include:

- Implementation of target CPU structure (e.g. registers, flags, etc.);
- Implementation of correct instruction timing, making the CPU being real-time;
- Whether it is effective to implement other devices in OpenCL, too;
- The communication among virtual devices (implementing signals, interrupts, input/output instructions, etc.);
- Implementation of operating memory;
- Choose or design new emulation technique that would be of better performance like decode-and-dispatch interpreter.

This section discusses the structure and implementation of a CPU emulator on a host GPU.

The general structure of a CPU emulator is represented by a 'context' defined as a data structure in used programming language. The context collect all internal memories of the CPU, including its special internal state. The memories are called registers; the internal state is represented by flags, etc. Now it would be appropriate to state that generally every CPU emulator imitates the ISA (Instruction Set Architecture) of a CPU. Therefore, the structure basically contains all necessary CPU storage structures that are visible from programmer's point of view only. The storage structures are implemented as single variables (or array of variables) that store values of appropriate data type. Generally, it is possible to encode each data type, when constructing some representation of the type that is of another data type. The context then can be understood as an instance of the structure in a time moment (the variables will be modified during the emulation).

Within the GPU, the idea of 'context' structure holds, too. The context is implemented as a memory region (memory object), logically divided into well-defined parts (Fig. 3). The region can be considered as an array with several equal-sized cells. The cell may be of any data type, however restricted to OpenCL types. The region size is computed as $sizeof(cell\_type) \cdot number\_of\_cells$.

Registers in a processor are tiny but high-speed memory cells that store binary values. Registers within GPU language are implemented as variables of some data type that covers the whole register size. Mostly, their purpose allows the same data type for all registers. Allocating and transferring register values between the host and OpenCL device one by one can decrease the performance, therefore better way is to use single memory array for all registers. The ISA structure of a virtual CPU does not consist of registers only, but include (for example):

- General-purpose registers;
- Program counter;
- CPU running state (supporting at least running and stopped values);
- Flags.

The structure is then called a *context*. The idea is that both the emulation algorithm and context are stored in GPU, in global memory (according to OpenCL GPU architecture). Another option is to use of (faster) local memory. However, host-GPU transfer goes through global memory and therefore it depends on how often the context is transferred between host CPU and GPU. Also, when the emulator uses several
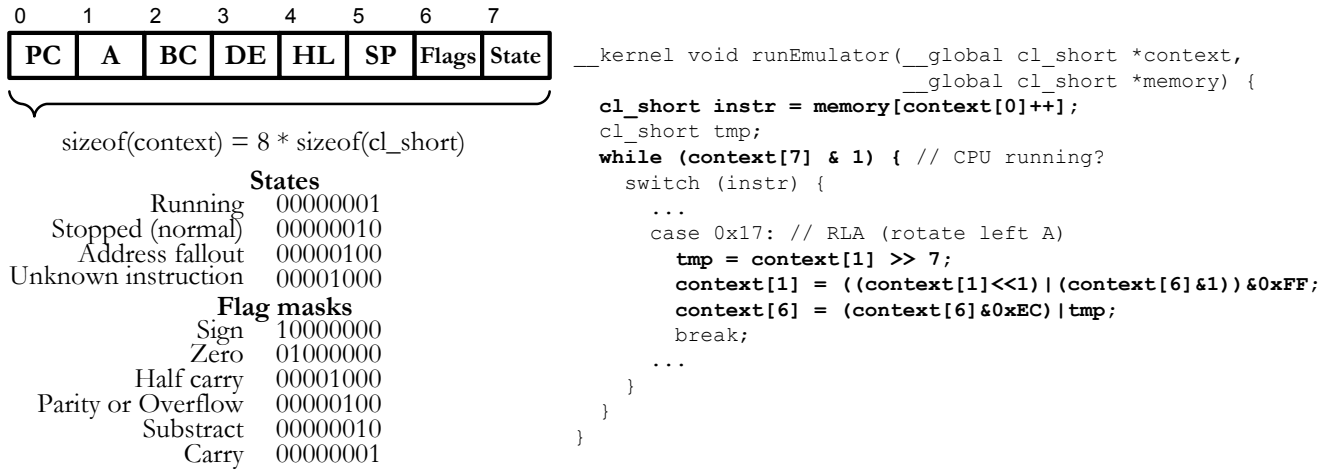
0	1	2	3	4	5	6	7
PC	A	BC	DE	HL	SP	Flags	State

$$\text{sizeof(context)} = 8 * \text{sizeof(cl\_short)}$$

States

Running	00000001
Stopped (normal)	00000010
Address fallout	00000100
Unknown instruction	00001000

Flag masks

Sign	10000000
Zero	01000000
Half carry	00001000
Parity or Overflow	00000100
Substract	00000010
Carry	00000001

```
__kernel void runEmulator(__global cl_short *context,
                          __global cl_short *memory) {
cl_short instr = memory[context[0]++];
cl_short tmp;
while (context[7] & 1) { // CPU running?
  switch (instr) {
    ...
    case 0x17: // RLA (rotate left A)
      tmp = context[1] >> 7;
      context[1] = ((context[1]<<1)|(context[6]&1))&0xFF;
      context[6] = (context[6]&0xEC)|tmp;
      break;
    ...
  }
}
}
```

Fig. 3. The example CPU context used in fictious Z80 CPU emulator

streaming kernels that would not be logically groupped into the same work-group, they cannot share data stored in local memory.

VII. CONCLUSION

The usage of simulators and emulators is indisputable – it is the validation model. The advantages of full system simulation/emulation are several. Software development can take place before the hardware is ready, therefore validation is faster. Validation is less costly and faster because many engineers can run validation tests on a PC instead of sharing a few HW prototypes. Some things can be done with simulation that can hardly be done with hardware, like verifying correct hardware initialization, simulating defective hardware, internal observations, etc. And the last thing, simulation tools can be connected with formal methods tools.

Stream processing and GPGPU brought a new potential of making high-performance software. They allow it by limited form of parallelism. Applications thus are divided into two parts – host part and client part, where host part is implemented in any classic programming language as we know today and it is executed on CPU. Client part is written in specialized stream language which extends the language of parallelism and on the other side, has some restrictions. The client part is executed on GPU or another compatible device, which disburdens the CPU of computational work that can be parallelized.

This work points at utilizing this new approach in another direction, as was suggested. The intent is to develop simulators/emulators of CPUs that will run as client parts of the application, on a GPU. However, some restrictions are put to the emulator, caused by restrictions of a streaming language. The last chapter proves that despite of all issues, it is worth to study and discover new possibilities of this approach.

There are still some open issues that were not analyzed, but they require more research. The most important are:

- Preservation of the CPU speed – cycle timing accuracy;
- Device emulation;
- Effective communication between devices and CPU.

ACKNOWLEDGMENT

REFERENCES

[1] J. Smith and R. Nair, *Virtual Machines: Versatile Platforms for Systems and Processes*, 1st ed. Morgan Kaufmann, 2005.

[2] D. A. Patterson and J. L. Henessy, *Computer organisation and design*, 4th ed. Morgan Kaufmann, 2008.

[3] V. Barrio, "Study of the techniques for emulation programming," 2001, available at http://personals.ac.upc.edu/vmoya/docs/emuprog.pdf. [Online]. Available: http://personals.ac.upc.edu/vmoya/docs/emuprog.pdf

[4] S. Hudák, *Machine oriented languages (in Slovak)*. FEI of Košice, 2003.

[5] S. Venkatasubramanian, "The graphics card as a streaming computer," *CoRR*, vol. cs.GR/0310002, 2003.

[6] U. Kapasi, W. J. Dally, S. Rixner, J. D. Owens, and B. Khailany, "The imagine stream processor," in *Proceedings 2002 IEEE International Conference on Computer Design*, sep 2002, pp. 282–288.

[7] B. Khailany, W. J. Dally, S. Rixner, U. J. Kapasi, J. D. Owens, and B. Towles, "Exploring the vlsi scalability of stream processors," in *Proceedings of the 9th International Symposium on High-Performance Computer Architecture*, ser. HPCA '03. Washington, DC, USA: IEEE Computer Society, 2003, p. 153. [Online]. Available: http://portal.acm.org/citation.cfm?id=822080.822826

[8] *NVIDIA OpenCL programming guide for the CUDA architecture. Version 3.2*, 2010, available at url: http://developer.download.nvidia.com/compute/cuda/3_2_prod/toolkit/docs/OpenCL_Programming_Guide.pdf.

[9] *The OpenCL Specification*, 2010, revision 36. [Online]. Available: http://www.khronos.org/registry/cl/specs/opencl-1.1.pdf

[10] "Direct compute lecture series," 2010. [Online]. Available: http://channel9.msdn.com/tags/DirectCompute-Lecture-Series/

[11] *AMD Accelerated Parallel Processing OpenCL*, 2010. [Online]. Available: http://developer.amd.com/gpu/ATIStreamSDK/assets/AMD_Accelerated_Parallel_Processing_OpenCL_Programming_Guide.pdf

[12] "Next generation cuda architecture," 2011. [Online]. Available: http://www.nvidia.com/object/fermi_architecture.html

Steps against breaking the safety of wireless network systems

Krekáň Ján

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

krekan@johnykrekan.com

Abstract

The main purpose of this paper is to demonstrate the security weaknesses found in existing wireless networks. It shows the principles of different security algorithms, and sketches up the ways how the attacker tries to break them. Securing the network against unauthorized access is a very significant part of the subject, but securing the network against dangerous activities of already authorized clients is very important too. We also try to discuss several methods and proposals for implementing a new model of monitoring the wireless networks. The goal of this monitoring is to allow system administrators to limit access to the network for users who are trying to break the security rules given by the network administrator.

Keywords

Authentication, authorisation, hack, monitoring, WEP, WPA

I. INTRODUCTION

First decade of the 21th century is characterized by rapid development of wireless networking. Mobile phones, Wireless networking, devices connected to each other using Bluetooth internet and several other technologies. Developers of those very comfortable and very popular solutions often underestimate the need for applying proper security measurements to protect data which travel in air. Securing and protecting wireless networks is very important in today's world where value of information is crucial for all of us. The security of wireless network is very important because these networks, when improperly configured are easy targets for attackers, who can break into those networks, steal information, or cause damages to organizations/companies which are using them. Since electromagnetic signals often spreads more far away than is estimated by the designer of a network topology for an organization, data transmitted from this organization's workstations can get to third parties. Several researches show that experienced computer user can get to data of third party organizations with just passive monitoring the air frequencies. Main goal of this article is to show necessity of implementing proper security models in wireless networking to prevent their eavesdropping. confidence of wireless networks and their proper configuration is important also because there are plenty of freeware software tools available on the internet which can be used to attack the wireless network or passively monitor the network, which is even more dangerous, because when network is monitored in this way, legitimate network users can not have any clues about the fact that their network is illegally monitored.

IEEE802.11 Overview

Organisation, structures, low-level communication protocols and security models are described in documents issued by IEEE working group. The letter appended after the name IEEE802.11 identifies more deeply the current issue solved by the working group – like frequency spectrum or safety standards etc. Standard 802.11b occupies the frequency range 2.4-2.4835 GHz and can operate up to 11 Mbps, although it can reduce its throughput down to 5.5, 2 or 1 Mbps according to signal quality. The 802.11a standard, which is not compatible with b/g hardware, can operate on 5.15-5.35 GHz and 5.725-5.825 GHz frequency ranges, having theoretical throughput up to 54 Mbps. Its disadvantage in practical applications is that the devices require external antennas positioned so they have a direct visibility to the access point. On the other hand they can communicate on larger distances. Finally 802.11g introduces 14 channels in the same spectrum as 802.11b, which means 802.11g hardware has a backward compatibility. The newest 802.11n standard introduces several new approaches to increase the overall performance. They do not come up with any new security improvement and this paper will not discuss them more deeply.

Authentication

Authentication is the process of proving identity of a station to another station or AP. In the open system authentication, all stations are authenticated without any checking. A station A sends an Authentication management frame that contains the identity of A, to station B. Station B replies with a frame that indicates recognition, addressed to A. In the closed network architecture, the stations must know the SSID of the AP in order to connect to the AP. The shared key authentication uses a standard challenge and response along with a shared secret key.

Association

Data can be exchanged between the station and AP only after a station is associated with an AP in the infrastructure mode or with another station in the ad hoc mode. All the APs transmit Beacon frames a few times each second that contain the SSID, time, capabilities, supported rates, and other information. Stations can chose to associate with an AP based on the signal

strength etc. of each AP. Stations can have a null SSID that is considered to match all SSIDs.

The association is a two-step process. A station that is currently unauthenticated and unassociated listens for Beacon frames. The station selects a BSS to join. The station and the AP mutually authenticate themselves by exchanging Authentication management frames. The client is now authenticated, but unassociated. In the second step, the station sends an Association Request frame, to which the AP responds with an Association Response frame that includes an Association ID to the station. The station is now authenticated and associated.
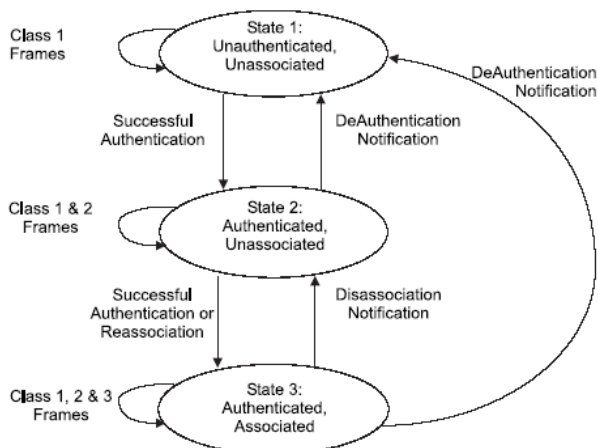

Fig. 1 – Authentication and Association

A station can be authenticated with several APs at the same time, but associated with at most one AP at any time. Association implies authentication. There is no state where a station is associated but not authenticated.

Current status of security in wireless networks

Experiments and exploration which took place in Kosice and other cities in Slovakia showed that the level of confidence and privacy in wireless networks is quite poor. When we assume that wireless 802.11 technologies are commonly used more than 5 years, it should be expected that security in those networks could be on a quite high level. Unfortunately, the results are very poor. At least a one half of monitored networks were not sufficiently secured. At least one quarter of all networks were not secured at all (they allowed anyone to get online trough them). Even more dangerous is the fact, that most of wireless nodes of big internet service providers are still unsecured (some nodes only verify the MAC address of the client which cannot be considered satisfactory).

Date of testing	9. 12. 2010		11. 5. 2010	
Networks found	73	100,00%	60	100,00%
Unprotected networks	15	20,54%	10	16,67%
Unsecured (WEP)	10	13,70%	6	10,00%
Protected networks	48	65,76%	44	73,33%

Tab 1.: Safety check monitoring in different seasons

Monitoring wireless networks

Monitoring of wireless networks is much easier in comparison to standard, wired networks. When a person wants to monitor

wireless LAN, first requirement is that the monitoring device must be in the range where this network can be heart. Secondly the monitoring machine must be equipped with suitable wifi adapter which must support monitoring in passive mode. Passive mode means that during process of monitoring, adapter is not connected to any network. In first step scan of all wifi channels is performed. After scanning, monitoring person chooses channel where is transmitting network which is going to be monitored. Monitoring can be performed on channel, or network basis. When performed on channel basis, all other (if any) nodes broadcasting on the channel where our targeted network broadcasts, are monitored as well. When monitored on a network basis, all software for monitoring Wifi networks found on the internet allows to specify filter based on BSSID (MAC address of the node), which allows to capture only data which belong to our network, nothing more.

When the data are collected data, analytical tools can be applied. By analysing the captured data it is possible to extract all not encrypted data from them, like e-mails, plain text passwords, record unencrypted VOIP conversations. All this can be performed absolutely invisible to the users of the monitored network. Since monitoring side does not transmit any data, it cannot be tracked. This is the main reason why using strong and properly implemented cipher standards are crucial for secure wireless networking.

II. ANALYSIS OF COMMON SECURITY MODELS USED IN WIRELESS NETWORKING.

MAC address controlling

This security measurement is often used to prevent unauthorized clients from connecting to the network. Every authorized user has its MAC address saved in a database of allowed addresses. When client attempts to connect to the network, his MAC is compared to the table of allowed MAC addresses, usually this table is saved in router. When the MAC of connecting client is found in this table, user is allowed to connect to the node, otherwise access is refused.
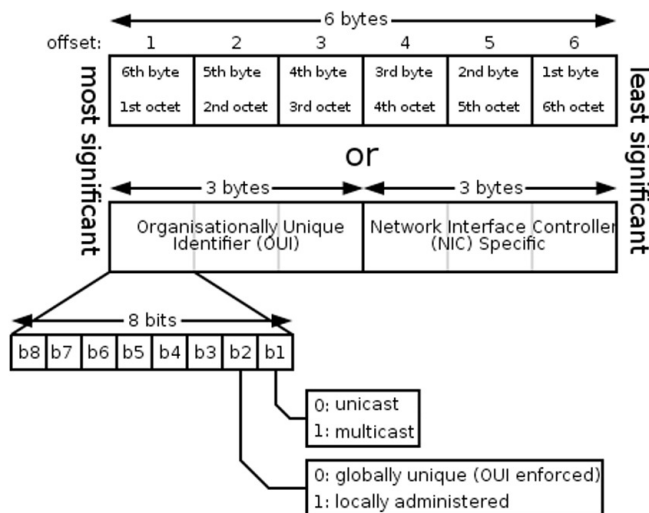

Fig 2 - MAC 48 address format

This model is often used without any ciphering of transferred data in the good will that it is sufficient. This model of security is totally inadequate, because when we can monitor the network as described in previous paragraph, we can easily

see MAC of every client associated with the node. When we have address of a client, it is very easy to spoof our MAC address to the one allowed. Next the real client is disconnected using active attacks like node re-association and the attacker can connect to the network with identity of previously disconnected user.

WEP

Wired Equivalent Privacy (WEP) is a protocol for encrypting wirelessly transmitted packets on IEEE 802.11 networks. In a WEP protected network, all packets are encrypted using the stream cipher RC4 under a common key, the root key. The root key is shared by all radio stations. A successful recovery of this key gives an attacker full access to the network. Although known to be insecure and superseded by Wi-Fi Protected Access (WPA), this protocol is still in widespread use almost 6 years after practical key recovery attacks were found against it.
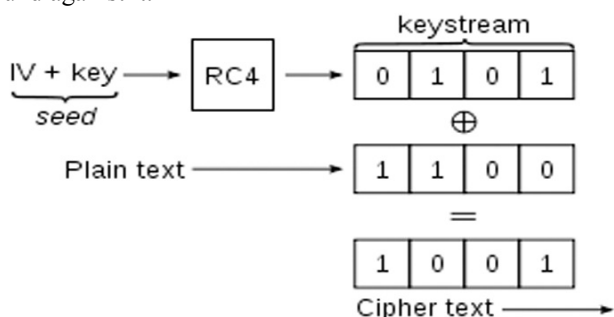

Fig. 3 – encryption algorithm WEP

Breaking WEP

Breaking of WEP protected network can be realized using two fundamental methods.

1. Passively monitor the network in order to collect necessary data which will be later used to recover the key used to protect the attacked network. This is absolutely untraceable attack since no data is broadcasted to the network. Main disadvantage of this attack is that it requires more time. The time needed to crack the WEP key using passive attack depends mainly on the network utilization during monitoring. The golden rule in WEP breaking is "more packets collected the better". When a network is not heavily used collecting sufficient number of data packets can take a few days. On a network which is frequently used it can take a few hours.

2. Active attacks. In WLANs with low network traffic levels, collecting thousands of unique packets necessary for key recovery (at least 40,000 for a 64-bit key) can be very time consuming. Tools like Aireplay (part of Aircrack-ng suite) or Commview for wifi can induce a higher traffic level by sending encrypted ARP Request packets. The WLAN would then reply with ARP Response packets, thus generating additional traffic. While the original ARP Request packets being replayed cannot contribute to the pool of the packets needed for key recovery (they all have exactly the same contents, while key recovery requires unique packets), the ARP Response packets are different from each other, as a new Initialization Vector is used for every new packet. Linux tools such as the above mentioned Aireplay can automate this process and it means that the software itself can find the proper kind of packets to inject. For WEP attack three tools are used. Airodump for capturing WLAN data, Aireplay for traffic generation and finally aircrack-ing for analysis of

captured data. When enough data is present, aircrack-ing will recover the original WEP key. The experience shows that attacking a wifi network using active methods can crack a 128 bit network in 5 minutes under good conditions. Summarising, it is sad to see that WEP is in still common use today, 6 years after it has been attacked successfully and marked as insecure.

Wireless Protected Access (WPA)

First, the WPA standard can be divided into two different security models: WPA Personal and WPA Enterprise or simply WPA.

In personal mode, WPA uses a single key for authenticating all clients that know the key. WPA uses robust encryption, per station keys, periodical rekeying and other measurements which prevent attackers from misusing the protected network. Generally can be said, that when administrator of a network chooses strong password (not word from dictionary or some derivation of it), breaking such password is almost impossible using computational power available to common attackers in these days. In 2009 was discovered that WPA1 personal which is using the temporal key integrity protocol TKIP, is not so secure as everybody thought. Scientists were able to break it in a time frame around 15 minutes. The WPA2 CCMP with AES standard used for data encryption remains unbroken also in the time of writing of this article. First advantage of WPA personal security model is that minimum password length was set to 8 characters. This prevents people from choosing short passwords (40 bit WEP allowed password which was composed only from 5 alphanumeric characters). Second positive property of WPA personal is the fact that password attack on this encryption scheme is very slow. On modern quad core Intel processor with one additional graphics card used for attack, this machine can test only 4000 passwords per second. That is absolutely not enough when one wants to test large amounts of passwords. Said differently, when attack with large dictionary (dictionary with more than 50000000 words) fails, the tested password can be considered secure. As was previously said since minimum password length is 8 characters, brute force attack is absolutely meaningless.

Another attack on WPA personal can be done using pre-computed rainbow tables. Main principle is as follows:
WPA rainbow tables must be pre-computed for specific SSID. Since many nodes use default names like linksys, airlive, and netgear... one can pre-compute table based on a large wordlist and make them available to the internet. Main advantage of using tables is speed. Trying a large table of several million words in size takes only a short period of time compared to a regular dictionary attack. Concluding this we can say that WPA2 personal when configured properly with strong authorization passphrase can be considered unbreakable.

WPA enterprise, simply called as WPA, is a robust standard used for authentication in large companies, schools where multiple clients connects to network. Every user has assigned user name (login) and password which are checked using a radius server. This approach is recognised as secured as there has been no success in breaking WPA enterprise yet. There is only one option left, hacking the radius server from outside, but this is not the subject of this papers.

III. SECURITY IMPROVEMENT PROPOSALS

To make the Wireless network more reliable and safe against the attacks from inside, we should eliminate all the suspicious activities. This task shall be divided in two parts – users verification and authentication, and monitoring the users already associated to the network.

Authentication is based on choosing the correct and secured password. The purpose of verification is to verify that the person who entered the password is eligible to access as she or he knows some additional information such as login (in case of WPA-Enterprise) or the computer has a correct MAC address associated with the registered name and password.

The main goal of monitoring is to protect people already connected to the network against the attackers hidden among them. We must be careful and find the most possible about port monitoring, active sniffing (APR), port scanning, packet flooding and similar activities. They are the critical signals pointing to illegal user actions.

In case of such an act detection the network rooter must close connection to the user and shall not allow a new connect until the administrators will not solve the situation and raise the ban.

The attacker can change his MAC address, but if the network is WPA-Enterprise where MAC is interconnected with login and password, this will not let him to keep on his activities. WPA-PSK is not so strong protected and this change can give a chance to continue, if he has MAC addresses of other common users to misuse. As to get these addresses is not so easy, we can say even WPA-PSK gives a reasonable security level to the network.

We shall also scan the users after their associating with the wireless network to gather some important information about their computers that act like human fingerprints. These data are associated with the client, they are unique to each computer, and they are hard to fake or change. Here we can include the operating system specific information such as OS version, version of services running and other data possible to find by a legal network scanners. If a user changes them without previously breaking the network rules, he will not need to inform the administrator and they will be accepted as normal. If the fingerprint change happens after the rules violation, the user shall be still unaccepted.

Improvements on user's side

Using WEP for encryption must be avoided as it is really easy to break. Also WEP has never been designed to provide a complete security solution for wireless networks, only a level of privacy equivalent to wired LAN. Traffic between wired LAN and wireless network must be limited to minimum and before passing any data authentication must be required. This recommendation is based on the fact that wireless network faces much stronger security challenges and if the attacker succeeds in his attempt, the wired part can be vulnerable as well.

The network SSID should include the names or any descriptions. This option is recommended as many hackers choose their targets precisely for these reasons. The attack and the data stolen can be the trophy presented to their community, or the network can be targeted as a revenge for any, real or imaginary, social injury. It is very welcome if the SSID is not transmitted at all (hidden).

If the access point has an option to disable beacon packets, it should be selected. The network does not broadcast the packets that can also help the attacker to identify the target.

Changing the encryption keys periodically would not prevent the compromise of WEP keys, if WEP security is used; anyway it slows the new attack down and shows that the network would never stay compromised indefinitely. Fortunately, the process of key management can be automated according to the 802.11i Task Group.

DHCP shall not be used on WLANS as well. To access host at a targeted site, an attacker needs to obtain a valid IP address and subnet mask on the WLAN. Without DHCP, identifying IP address requires sniffing the traffic and reviewing the captured data. This means that the static IP addresses are one more deterrent that may discourage the attacker.

Replacement of Advanced Encryption Standard (AES) instead of Data Encryption Standard (DES) is also recommended. It is described in 802.11i Task Group and government officials in several countries already issued instructions to follow the process to enhance the network security.

IV. FINAL WORDS

Testing of wireless networks in the area of Kosice showed that their security is absolutely inadequate. What is worse is the fact that those networks offer a good opportunity for making cyber crime. Other effect of this is the fact that tracking down criminals who used unsecured wifi networks for their crime is almost impossible when they perform their attack without mistakes. This is also one of the reasons to take this topic for the serious research and implementation, because as it can be easily proved believe, well secured networks can make the whole life more secure.

V. ACKNOWLEDGEMENTS

VI. REFERENCES

[1] Stuart J. Kerry, "Chair of IEEE802.11 Responds to WEP Security Flaws", Feb. 15, 2001. Available at http://slashdot.org/articles/01/02/15/1745204.shtm

[2] http:// standards.ieee.org/about/get/802.11.html

[3] IDEFENSE Intelligence Report ID#106452, Dec. 6, 2002

[4] Arbaugh, William, Narendar Shankar and Justin Wan. "Your 802.11 Wireless Network has No Clothes", Available at http://downloads.securityfocus.com/library/wireless.pdf

[5] Borisov, Nikita, Ian Goldberg and David Wagner. "Intercepting Mobile Communications: The Insecurity of 802.11" Available at http://www.isaac.cs.berkley.edu/isaac/wep-draft.pdf

Stochastic Update of Backpropagation Through Time Algorithm

^{1}*Juraj KOŠČÁK,* ^{2}*Rudolf JAKŠA,* ^{3}*Peter Sinčák*

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

^{1}jurajkoscak@gmail.com, ^{2}jaksa@neuron.tuke.sk, ^{3}peter.sincak@tuke.sk

Abstract—s **We will examine various modifications of back-propagation through time algorithm (BPTT) done by stochastic update in the artificial recurrent neural networks (ARCNN). The general introduction was provided by Salvetti and Wilamowski in 1994 in order to improve probability of convergence and speed of convergence. This update method has also one another quality, its implementation is simple for arbitrary network topology. In stochastic update scenario, constant number of weights/neurons is randomly selected and updated. This is in contrast to classical ordered update, where always all weights/neurons are updated. Stochastic update is suitable to replace classical ordered update without any penalty on implementation complexity and with good chance without penalty on quality of convergence. We have provided first experiments with stochastic modification on backpropagation algorithm (BP) used for artificial feed-forward neural network (AFFNN) in detail described in the article [1] published on conference in Barcelona. We will present experiment results on simple toy-task data of postponed and modified signal as a verification of our implementation of different algorithm modifications.**

Keywords—**artificial recurrent neural networks (ARCNN), backpropagation algorithm (BP), backpropagation through time (BPTT), stochastic update.**

I. Introduction

The stochastic weight update was introduced by Salvetti and Wilamowski in 1994 [2] in order to improve probability of convergence and speed of convergence of the backpropagation algorithm. Besides the stochastic weight update, they examined another two stochastic methods: random pattern selection and randomized learning rate. On the XOR problem they demonstrated significant improvement in the learning speed and probability of convergence for every one from these methods, especially for randomized learning rate. The backpropagation algorithm (Werbos, 1974; Rumelhart, Mc-Clealland, 1986) is one of most used learning algorithms today. Plain vanilla implementation and momentum implementation (Silva, Almeida, 1990) [3] are the predominant implementations on feed-forward topologies. Multilayer perceptron (MLP) is most used feed-forward topology, it is a layered architecture with fully interconnected layers. The BPTT [4] is most used backpropagation variant for recurrent topologies. It is well suited algorithm for recurrent topologies with full connectivity. Although the BPTT implementation is not much more complex compare to feed-forward vanilla backpropagation, the feed-forward time-delay networks are preferably used for time-related problems. Time-delay networks are of the MLP design, so layered topology is most used topology with the backpropagation algorithm. Sparse topologies became recently popular

with the recurrent echo state networks (ESN) (Jaeger, 2001) [5]. However these are usually not based on backpropagation algorithm. ESN networks are based on so called reservoir of randomly

II. Stochastic Weight Update

Salvetti and Wilamowski did incorporate stochastic processes into backpropagation algorithm in order to improve convergence - to avoid to get stuck in a local minima on error surface [2]. Stochastic Weight Update was one from several ways how to do it. During learning, maybe it is not necessary to update all weights. Not all of them have the same influence on the network error. Some of them are more important, maybe have higher values, some of them are close to zero, maybe their influence is lower, or none. If we don't update all weights, we can also save some computation time and thus maybe speed up the learning process.
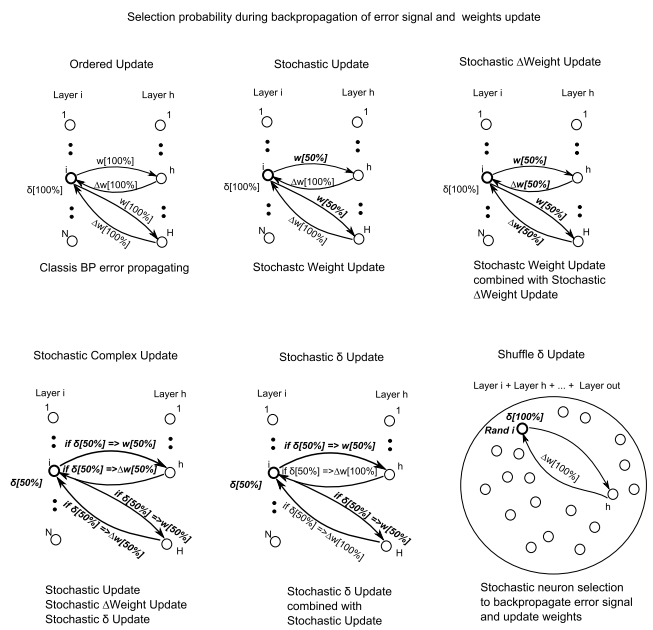

Fig. 1. Visualisation of implemented algorithms using stochastic factor. Each Stochastic approach uses 50% ratio that the selected weight/δ is going to be updated. The stochasticity is implemented on the different levels of learning algorithm as they are describe on the Figure.

1) Ordered Update: Classic BPTT algorithm, where all weights are update in one iteration. Error signal is computed for all neurons without any preferences.
2) Stochastic Update: Modified BPTT algorithm, where not all weights are updated in the same iteration. The chance

of weight update is controlled by r, where $0 \leq r \leq 100$, which represents the probability of weight update in percentage. The weight is selected for update if the generated number is $rand \leq r$. Otherwise the weight is not going to be updated. The ARCNN keeps this way information from the past.

3) Stochastic Δ Weight Update: This approach combines the Stochastic Update with the stochastic selection of δweights, that are going to update. After the back propagated error signal is computed, the δ weights are going to be update. The stochastic approach is the same as with Stochastic Update. δ weight is updated if the $rand \leq r$.

4) Stochastic δ Update: Combines Stochastic Update with stochastic selection of δ Update. Stochastic factor is the same, as in the rest of stochastic modifications. The δ is selected to be update with probability. This selective updating of δ cases, that also here some old δ from $t-1, ..., t-x$ (x - not known the past, that haven't been update yet) acts in the learning process and have influence on the whole process.

5) Stochastic Complex Update: The stochastic factor of probability is the same as in previous stochastic methods. It combines Stochastic Update, Stochastic δWeight Update and Stochastic δ Update.

6) Shuffle δ Update: This approach simplifies the implementation BPTT algorithm. It is more simple because you do not have to care about layers and the order of the neurons. Simply one neuron is randomly selected from the "pool of neurons", error signal is computed and all incoming weights are updated. The "pool of neurons" represents all neurons in all layers, that contribute on δ calculation.

If we want to increase weight selection chance over 100% (all weights will be updated) we can add another loop with parameter, which control how many times this random selection should be done. If set to 1 it is equal to the loop missing.

III. Experiments

To test the Stochastic Weight Update we chosen the modified and postponed signal that proves the correct implementation of the ARCNN and also provides an easy to prove the stochastic update modifications of the BPTT algorithm - a toy-task benchmark (see Fig. 2).
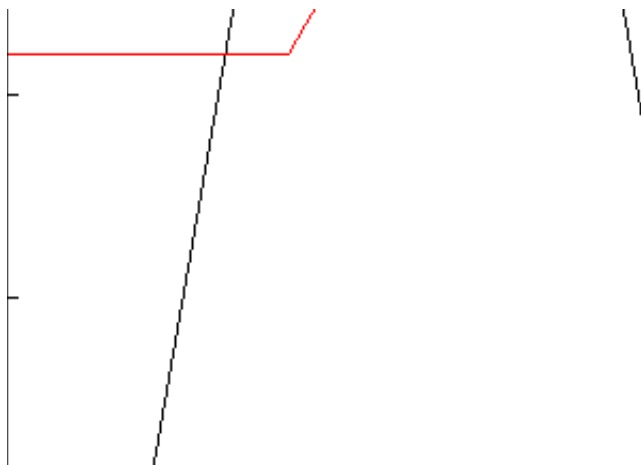

Fig. 2.   Moved and modified signal used for learning ARCNN.

We wanted to check that the Stochastic Weight Update will not reduce quality of backpropagation learning, and also to explore the influence of ratio of weights updated per pattern on the results. The results are compared by the lowest least mean square (LMS) error.

We will provide 2 experiments, where we compare different types of stochastic implementation, as described on the Figure II, and ordered classic implementation of BPTT algorithm on ARCNN. We use the same number of neurons, the same sigmoid function with the steepness λ and the same learning rate γ.

First experiment consist of the 5 ARCNN using different modification of stochastic update (see Figure II). Topology of all ARCNNs is based on the best results of the ordered update. The Figure 3 represents the results of the first experiment, error is presented on Figure 6 and detailed settings of ARCNNs are in the Table I. The experiment was orientated to achieve the better results using one of the stochastic modification compared to the origin update of BPTT implementation on the ARCNN.
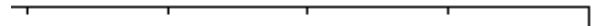

Fig. 3.   Experiment output of the six ARCNNs learned by the configuration based on the best of the ordered update of BPTT.
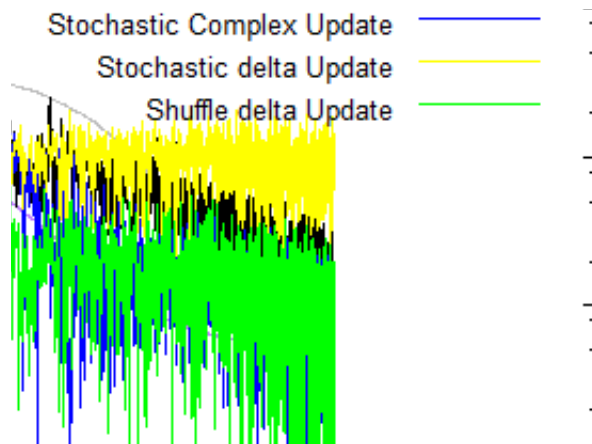

Fig. 4.   LMS error of the experiment based on the best of the ordered update of BPTT.

Second experiment is little bit different from the 1st experiment. Experiment uses also α momentum to get better result and do not rely just on the algorithm ability to deal with signal. The conclusion of the experiment is, that the α momentum can be used as a final touch to improve learning of ARCNN.

TABLE I
TO ACHIEVE BETTER RESULTS THEN THE ORDERED UPDATE

Method	Ordered Update	Stochastic Update	Stochastic ΔWeight Update	Stochastic Complex Update	Stochastic δ Update	Shuffle δ Update
Selection Ratio	100%	50%	50%	50%	50%	50%
γ	0.03	0.03	0.03	0.03	0.03	0.03
λ	0.65	0.65	0.65	0.65	0.65	0.65
α momentum	0.0	0.0	0.0	0.0	0.0	0.0
Unfolding history T	1	1	1	1	1	1
Topology	1-15-1	1-15-1	1-15-1	1-15-1	1-15-1	1-15-1
Adaptaptation in one cycle	1	1	1	1	1	1
Iterations	1500	1500	1500	1500	1500	1500
LMS error	0,00572	0,01063	0,00449	0,0106	0,197	0,00071

Table represents the settings of the ARCNNs for the 1st experiment - achieve better results then the ordered update of BPTT. All ARCNNs have same settings. The Shuffle δ Update modification was able to achieve higher accuracy.

TABLE II
USING α MOMENTUM FOR ARCNN

Method	Ordered Update	Stochastic Update	Stochastic ΔWeight Update	Stochastic Complex Update	Stochastic δ Update	Shuffle δ Update
Selection Ratio	100%	50%	50%	50%	50%	50%
γ	0.03	0.03	0.03	0.03	0.03	0.03
λ	0.65	0.65	0.65	0.65	0.65	0.65
α momentum	0.95	0.95	0.95	0.95	0.95	0.95
Unfolding history T	1	1	1	1	1	1
Topology	1-15-1	1-15-1	1-15-1	1-15-1	1-15-1	1-15-1
Adaptaptation in one cycle	1	1	1	1	1	1
Iterations	2500	2500	2500	2500	2500	2500
LMS error	0,00107	0,00098	0,00145	0,00166	0,00133	0,00141

Experiment results table, where was used α momentum to achieve higher accuracy. The best results were achieved by the Stochastic Update modification.

Figure 5 represents the results of the experiment, error is presented on Figure 6 and detailed settings of ARCNNs are in presented the Table II.

Fig. 5. Output of the ARCNNs for the BPTT using the α momentum. The best results were achieved by the Stochastic Update modification (see Table I).
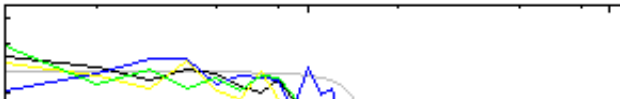
year 2000 till the end of February 2011. Dataset is split into two groups. Training dataset, which is provided to network in learning phase, and testing dataset, which is unknown part of dataset to ANN and is used to test the ability of ANN to predict the temperature and the ability of ANN to generalize the model hidden in the temperature profile.

REFERENCES

[1] J. Koščák, R. Jakša, and P. Sinčák, "Stochastic weight update in the backpropagation algorithm on feed-forward neural networks," in *Neural Networks (IJCNN), The 2010 International Joint Conference on*, 2010.

[2] A. Salvetti and B. M. Wilamowski, "Introducing stochastic processes within the backpropagation algorithm for improved convergence," in *AN-NIE'94 - Artificial Neural Networks in Engineering, (Intelligent Engineering Systems Through Artificial Neural Networks)*, C. Dagli, B. Fernández, J. Ghosh, and R. Kumara, Eds., vol. 4. ASME PRESS, New York, 1994, pp. 205–209.

[3] F. Silva and L. Almeida, "Acceleration technique for the backpropagation algorithm," in *Neural Networks EURASIP workshop*, 1990.

[4] P. Werbos, *The Roots of Backpropagation: From Ordered Derivation To Neural Networks and political forecasting*, February 1994.

[5] H. Jaeger, "Gmd report 148, the "echo state" approach to analysing and training recurrent neural networks," GMD - German National Research Institute for Computer Science, Tech. Rep., 2001.

Fig. 6. LMS error for the BPTT using the α momentum. The best results were achieved by the Stochastic Update modification (see Table I).

IV. CONCLUSION

Our results with the moved and modified signal confirm the original results of Salvetti and Wilamowski on XOR problem, experiments provided on the AFNN using the circle-square separation task and also on the ARCNN with modified and postponed signal - the Stochastic Weight Update performs well and in some cases can be superior to classical ordered update in backpropagation through time algorithm.

Complexity of implementation of this weight update is less dependent on the network topology than the classical ordered weight update is, while overall complexity of implementation is not much higher than the classical one. This can be utilized in BPTT implementations with non-standard, sparse, or random network topologies.

Our main goal is to provide the prediction of the temperature daily profile using the tested and implemented BPTT stochastic update due to the ability of ARCNN to use the history of previous steps stored inside. The dataset used for temperature prediction will consist of the measured temperature from the

Study of computer emulation algorithms

[1]*Marián MIŽIK*, [2]*Peter JAKUBČO*

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]marian.mizik@tuke.sk, [2]peter.jakubco@tuke.sk

Abstract—**This paper presents proposal of sequential algorithm used for computer emulation. We show, that the sequentiality is a natural consequence if control-flow processors are used. The algorithm works in a loop, where the CPU and the other computer components are emulated step by step. However, sequentiality might have negative impact for the quality of whole emulation. This problem is solved by the emuStudio platform, used as universal emulator. The solution depends on some assumptions that the emulated components have to fulfill..**

Keywords—**control flow, emulation, sequential algorithm**

I. INTRODUCTION

The main activity of running emulation is repeated execution of the same sequence of steps, because the work of real running computer can be viewed as a repeated sequence of activities [2]. A CPU (processor) fetches, decodes and executes instructions in infinite loop. The other computer components are directly or indirectly controlled by the processor. Their activities are independent, parallel and can perform in synchronous or asynchronous way with the work of the CPU.

The implementation of parallel and asynchronous activity of emulated component is more difficult, because the starting time of the activity is not predictable (and it should not be), and depends on events that fire in the component. The nature of the most host processors does not allow the true parallel work – if we want to run the component in true parallel with CPU, it must not to be executed on the same host processor (if the processor has only one physical core). Generally, the nature of control-flow processors (here belong all computers of von Neumann type) defines sequential execution of programs. Seeing that an emulator is also a program executed on the computer, its activity (even if repeated), has to be sequential.

II. UNIVERSAL EMULATION

Computer emulators (system virtual machines), besides processor emulation, often realize peripherals emulation as well. In such a way complete architecture is emulated, or more precisely its certain configuration. Computer architecture is system characteristic that integrate a function (usage) and the structure of its components [6]. As the best-known, wide-spread architecture is Princeton (well-known as von Neumann) architecture [1] that is used as implementation model modern computers, too. Some success achieved also Harvard architecture [7]. Computer configuration is a specific selection of architecture components. The structure and

function of the components has to satisfy requirements of the architecture. Most of computer emulators emulate just one such configuration. In a case of von Neumann architecture emulator it means that it is emulated just one type of processor, just one type of main store and particular peripherals, connected to a static structure. Such emulator can execute software that meets with emulated configuration.

Disadvantage of such approach is that for every computer configuration new emulator must be constructed. An interesting idea is therefore to universalize the emulator, so that just one emulator would be needed for emulating several configurations. Easier implementation is to realize the emulator to have static core (processor, main store, bus), and that only peripherals could change in the configuration.

Much more interesting idea is however to implement the emulator in a way that it would be possible to change every emulated component dynamically, i.e. the processor and main store, too.

In the present, some virtualization programs (e.g. VirtualBox, QEMU etc.) provide some ability to make configuration changes. Seeing that it cannot be spoken about emulation in true meaning, the processor it is not possible to change. They mostly offer only some minor changes, such as selection of components used (from a static list) and some modifications in their settings. Their complex interconnections cannot be changed dynamically. In the present, there don't exist many of such universal emulators.

More or less successful attempt for emulated hardware standardization was realized in the M.A.M.E. emulator [2] with purpose to save historic games and older game consoles. In the present, by European Union donation, it is prepared a project of similar purpose [3]. However, the ability of the user to choose emulated computer configuration dynamically, is lacked in mentioned emulators (KEEP emulator is actually in preparation phase).

III. EMULATION DESIGN

The work of the emulator will be described by a sequential algorithm. Inside the algorithm a scheduler needs to be implemented. The scheduler will allow the host processor time to individual emulated components, including emulated processor, in exactly defined order. The basic algorithm of the emulator's work, using a simple scheduler, is described in the pseudo-code shown in Fig. 1. In the

algorithm, a specific time interval is allotted to every emulated component always in the same sequence. The component can perform an activity in that time interval.

```
running = true;
(numberOfCycles, maxTime) =initTimeSyn();

while (running) {
    running = executeCPU(numberOfCycles);
    ints = generateInterrupts();
    emulateDevices(ints, maxTime);
    (numberOfCycles, maxTime) = timeSyn();
}
```

Fig. 1 Basic emulator algorithm (sequential)

The time interval is estimated by the emulator, and the estimation depends on required emulation speed. For the sake of clearness, lets imagine how a preemptive process scheduler works in an operating system. Each such scheduler allots fixed time interval to processes, allowing them to run in the time interval. When alloted time fires up, the operating system stops the running process and the next process is launched, with new allotted time interval.

Emulators work similarly, however when the time interval fires up, time synchronization must be performed, because it is not guaranteed that the components will be executed in the exact boundary of the time interval (sometimes the emulation can not be stopped right after fired time interval – it can result from the nature of the emulated component). Usually the time interval is estimated for one emulation iteration for all components together, and finally the time synchronization is performed. The time interval presented in Fig. 1 is divided into two parts:

• numberOfCycles – the maximal number of machine cycles of emulated processor that should be executed in single emulation step.

• maxTime – the maximal common time interval, in which all devices can perform activities in the single emulation step.

Besides the CPU and devices emulation, there are also external interrupts generated that come from devices. They are represented by the set of corresponding state variables in the CPU emulator. The variables are set only if certain conditions are accomplished. The service of the interrupts is realized outside of the CPU execution. The time synchronization, if needed, ensures the balance of the emulation speed, so the emulation speed would be as close to required speed as possible. The final emulation speed (desired) mostly come out from real parameters of emulated components.

IV. THE ALGORITHM PROBLEM AND ITS SOLUTION IN EMU STUDIO PLATFORM

The emuStudio emulation platform is a software developed as a result of researc at the Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice.

Design aim was to create a platform, that should be able to emulate various computer configurations, that are based on the architecture with von Neumann conception [3]. The algorithm presented in the previous section (Fig. 1) deals with a problem, however it has not to be always exposed. As was said above, the algorithm is sequential. It means that steps of the algorithm are executed in an ordered sequence. The executeCPU procedure interprets instruction flow and handles data flow. The input parameter of the procedure is the number of machine cycles of the CPU, during what this unit processes instructions and data. This number is usually big enough for an execution of hundreds instructions though (a single instruction consists of several machine cycles).

The problem is, that in a single call of the executeCPU procedure, there can be executed even several instructions working with the same device. As the device emulation procedure (emulateDevices) lies out of the executeCPU procedure, the evident become only the last such instruction. Therefore the device can not respond to the change immediately - the sequential algorithm forces the device to "wait" for its time. As an example, let's present a situation where the processor communicates with a graphic card (Fig. 2). The processor in one emulation step may tell the graphic card to draw a point onto the screen, two times at the same position, but with different colors. In the real computer the user would see the change, but he would not in the emulator, because the graphic card will draw only the last point that appears in the performed emulation step of one single processor.

There exist several possible solutions for the problem. The universal one is to use a queue that will store all changes that should be performed in a device, in the exact order. The queue would be internal part of the device. When the emulator allots the time to this device, the device, step by step, dequeues the changes and performs them. The device can own several queues though and their implementation can vary, depending on the way how the processor communicates with
the device. However, the disadvantage of such solution is a fact that the changes can "pulse" from the user's point of view. At first, the queue "fills up" - in that time the user does not see any changes. Then the queue "flushes" and the changes will appear gradually. If the filling/flushing time ratio is bigger than 1, long periods of inactivity will alternate with too quick changes. In the emuStudio emulation platform another solution is used, depending on the ability of the host operating system (or at least Java) to work with threads.

The execution of each input-output instruction causes immediate call of corresponding input/output of the emulated component. The identification of the component is ensured by the communication model [4] that assumes quick response from communicating components. If the response of the device will not be quick enough for the processor, the slow activity should be executed in an individual thread and then the execution should return back to the processor. In repeated access to the device, another thread will be created for performing the input/output operations. However there exist cases when this approach is not possible to realize, however

the cases exist only for the input operation (e.g. when processor needs the result immediately in order to complete current instruction execution).
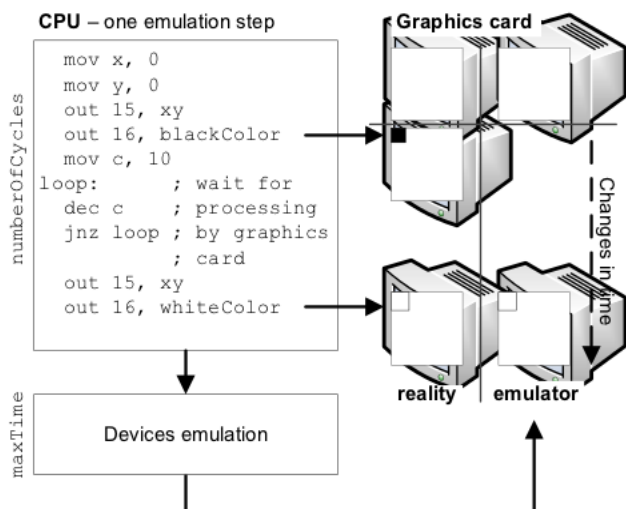

Fig. 2 Example of sequential emulation algorithm problem

In any case, the device owns the responsibility for potential synchronization requirements. Too slow responses of the devices can lead to system destabilization. There exists an empirical rule that says if the response will bound to one third of the time needed for numberOfCycles cycles execution, the latency influence to the system stability would be negligible. The pseudo-code implementation shown in Fig. 1 does not depend on used emulation technique. The proposed solution affects the executeCPU procedure in a way that the emulateDevices procedure (in modified form) will be nested in a certain way to the executeCPU procedure.

V. CONCLUSION

In the present time the emulation thematics is actual enough, mainly its special case virtualization. The principle of the technique is to emulate as few processor instructions as possible, and let the rest of them to execute natively on the host machine. However, the virtualization uses some emulation technique, too, and its similarity is indisputable with sequential emulation algorithm described in this paper. The paper focuses on the description of basic computer emulation algorithm. We can see its modification in any emulator and therefore it plays an important role.

By the nature of control-flow processors, that are controlled by the instruction flow in an exact order, the algorithm is sequential. Processor emulation in the standard algorithm is separated from the other computer components emulation, by what the individuality and independency of the components is emphasized. The components exchange the data to each other by certain communication interface. However this standard approach can have negative impact on emulation quality, what the paper shows in an example.

The emuStudio emulation platform, that is used mainly in the education process, solves the problem effectively. It combines the processor emulation with the emulation of the other components in a way that the communication between

the processor and the component is performed on demand, and quick response of the component is required.

REFERENCES

[1] P. Jakubčo and S. Šimoňák, "Emustudio – a plugin based emulation platform," Journal of Information, Control and Management Systems, vol. 13, no. 1, 2009.

[2] S. Šimoňák and P. Jakubčo, "Software based cpu emulation," Acta Electrotechnica et Informatica, vol. 8, no. 4, 2008.

[3] J. von Neumann, "First draft of a report on the edvac," Moore School of Electrical Engineering, University of Pennsylvania, Tech. Rep., 1945, available at http://qss.stanford.edu/~godfrey/vonNeumann/vnedvac.pdf

[4] P. Jakubčo and M. Domiter, "Standardization of computer emulation," in Applied Machine Intelligence and Informatics (SAMI), 2010 IEEE 8th International Symposium on Applied Machine Intelligence and Informatics, Herľany, Slovakia, 2010, pp. 221–224

[5] M. Jelšina 2002. Architektúry počítačových systémov: princípy, štruktúrna organizácia, funkcia. Košice: Elfa, 2002, 2nd edition. 567p. ISBN 80-89066-40-2

[6] L. Vokorokos et al.: 2008. "Digital Computer Principles" (in Slovak). Elfa, s.r.o., 322p. ISBN 978-80-8086-075-2

[7] L. Vokorokos, "Princípy architektúr počítačov riadených tokom údajov", Copycenter, spol. s.r.o., Košice, 2002, ISBN 80-7099-824-5

[8] L. Vokorokos, N. Ádám, J. Trelová "Pipeline system for processing of single operand data flow operators", International Conference on Applied Electrical Engineering and Informatics, Genoa, September 7-11, FEI TU, 2009, pp. 49-54, ISBN 978-80-553-0280-5

Subjective Video Quality Assessment by Gathering of Statistical Data

Martin KAPA, Juraj VALKUCAK

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

martin.kapa@cnl.sk, juraj.valkucak@cnl.sk

Abstract—With the importance of users' mobility the corresponding multimedia services need to be highly reliable to earn the trust of theirs customers. This is why we need precise assessment methods and frameworks to guarantee a specific level of user experience. This has partially been done with signal-to-noise ration (SNR), peak signal-to-noise ration (PSNR) or bit error rate (BER). However, the latest measurements have shown that they do not correlate well with quality perceived by an end-user. In order to achieve this, we propose a tool based on gathering of client-side statistical streaming data.

Keywords—video quality, subjective video quality assessment, video streaming.

I. INTRODUCTION

Multimedia streaming applications are becoming more and more popular. We have been able to see upward trend in providing of combined voice, video and data services. Video can be considered as a new feature in this scheme, but it importantly change the whole aspect of the way how we work, live, learn and communicate. With the importance of users mobility the corresponding multimedia services need to be highly reliable to earn the trust of theirs customers. This is why we need precise assessment methods and frameworks to guarantee a specific level of user experience. In order to achieve this goal we firstly need to consider which parameters have influence on degradation of user experience and quality of multimedia content. This has partially been done with signal-to-noise ration (SNR), peak signal-to-noise ration (PSNR) or bit error rate (BER). However, the latest measurements has shown that they do not correlate well with quality perceived by an end-user [1]. Therefore, concepts such as Quality of Service (QoS) and Quality of Experience (QoE) [2] [3] has been introduced. However, most of the current approaches are oriented to one specific video content type, specific application or scenario, which is not enough. Video quality metrics need to be more cross-content to provide better correlation with subjective ratings that are really important for appropriate decisions on a suitable optimization method for video streaming.

II. SUBJECTIVE VIDEO QUALITY ASSESSMENT STANDARDS

A. Absolute Category Rating (ACR)

This method is based on a category judgment where the particular test sequences are presented and rated independently on a category scale. The presentation of the video sequence is followed by the subjective questionnaire evaluating the quality of the shown sequences. The rating is done in five-level scale (5-excellent, 4-good, 3-fair, 2-poor, 1-bad). The necessary number of replications is obtained by repeating of the same test conditions at various points of time in the test.

B. Degradation Category Rating (DCR)

The degradation category rating is different to ACR by presenting sequences in pairs, where the first stimulus presented in each particular pair is the reference source. While the second stimulus is the same source sequence but presented through one of the systems under test. If a constant voting pattern time is used, then the voting time should be less or equal than 10s. However, the presentation time could be reduced or increased depends on the content of the test material. The rating is done in five-level scale (5-imperceptible, 4-perceptible but not annoying, 3-slightly annoying, 2-annoying, 1-very annoying).

C. Pair Comparison method (PC)

The assessment through the Pair Comparison method is realized by presenting a pair of identical test sequences. This sequence is firstly presented through one system under test and then through another system. The group of systems under tests (A, B, C, etc.) is generally combined in all range of possibilities. Therefore, all the pairs of sequences will be presented in both possible orders. If a constant voting pattern time is used, then the voting time should be less or equal than 10s. The presentation time should be about 10s long, but could be reduced or increased depends on the content of the test material.

D. Current research

The comparison between both the expert and non-expert viewers was measured in [4] by comparing the subjective video quality of several video sequences encoded with MPEG-2 and H.264 codecs at different bit rates. The results showed differences among the quality ratings, where experts provided lower ratings of video quality compared to non-expert viewers. The difference is generally small for high quality video, but become more pronounced for videos with high level of impairments. In [5] a subjective study was realized to assess the perceived quality of H.264 encoded video sequences transmitted over a wireless channel. Based on the results from the study, various objective quality assessment algorithms were evaluated using popular metrics. The used test reference sequences as well as the distorted sequences have been made available to the research community in case of future research. Four subjective video quality assessment experiments were

conducted in [6] using four different groups of viewers. The presented video sequences were all the same in all four experiments; however, the participants rated video quality using four different rating scales (5 point discrete scale, 9 point discrete scale, 5 point continuous scale and 11 point continuous scale). The results of these experiments indicate that most of the participants tend to distribute their ratings more evenly across the scale. This behaviour proves the existence of subjective response style, which is worth further investigation.

In [7] a novel subjective quality assessment method was proposed. It is suitable for comparing of various types of quality distortions. The method is called double stimulus adjustable fixed anchor (DSAQFA) and it allows a fast comparison among perceived qualities with a low amount of training necessary. The main purpose of DSAQFA is to compare the impact of source distortion on perceived quality in compare to channel distortion.

III. FUTURE ASSESSING TECHNIQUES

The view on user experience assessing techniques has dramatically changed. Current approaches do not provide accurate results, all of which are caused by inadequate designing process and misunderstanding of the QoE problematic. While designing future techniques/standards we have to bear in mind steps like Data Acquisition, Metric Selection and Results Analysis. To ensure particular level of user satisfaction every assessment platform needs to collect big amount of data. In the case of QoE those data should be acquired as close to the application users as possible to provide accurate estimation of perceived quality. On top of that, we have to think about what kind information we have to collect, where in our network we want to collect them and how we will realize the acquisition.

IV. PROPOSED SUBJECTIVE QUALITY ASSESSMENT

The general purpose of the tool is to provide monitoring of video playback on client side. Video is played in media player embedded in web browser using http streaming approach. Monitoring is useful to get client-side statistical video playback data, send them and store them on remote database server, when they can be evaluated. On this principle, quality of service can be easily determined and consequently improved while providing http video streaming. The tool is built on open source license software that is easily accessible, scalable and high-performance. The tool has distributed architecture, its functional parts work both in clients and server. The tool is composed of multiple components that as the whole provide desired functionality.

A. Functional topology and components

The system consists of several components - http server, web page located on the server with specific client-side scripting code, media plug-in in clients' web browsers and database server. Http server provides web page with link to the video to be streamed. Web page contains specific functions that exploit functionality of media plug-in. Media player plug-in is primarily used to playback streamed video on client side. Player is embedded in a web browser and processes the video that is being downloaded. Media player plug-in has been modified to provide needed functionality - to provide real-time video statistical data such as input bitrate, read packets, state of playback, frames per second, etc. Web page contains code

written in client-side web scripting language, so it runs on client side and handles the data gained from media plug-in. The gathered data can be sent to remote database where they are stored and third-party applications could access this data and evaluate them. The functional topology is shown in the [Fig. 1].
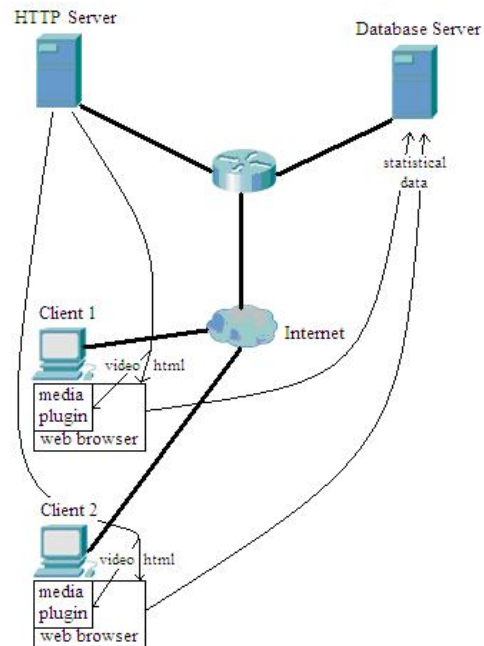

Fig. 1.   Functional topology

B. Implementation

To implement functionality of each component we chose different technologies. Firstly, we had to choose a open source media player that has a web browser plug-in module already built-in. Open source license was necessary because we needed to add our specific functionality to existing plug-in implementation and it would be very time consuming process if we started to develop media plug-in ourselves. We chose VLC media player from VideoLAN project that has also implementation of web browser plug-in called VLC multimedia plug-in. Secondly, we had to choose client-side web scripting language and our decision was Javascript, because VLC multimedia plug-in has support for Javascript already implemented.

1) VLC Multimedia Plug-in: VLC multimedia plug-in is a part of VLC media player that is developed by the VideoLAN Project, a project lead and composed by a team of volunteers, which believe in the power of open source when dealing with multimedia. VideoLAN produces free software for multimedia, released under the GNU General Public License, see http://www.videolan.org/videolan/. The source code of media player can be directly downloaded from http://www.videolan.org/vlc/download-sources.html. Its size is 22.8Mb. VLC multimedia plug-in in our case is used to playback streamed video from http server. Using its Javascript API with our extended implementation, we are able to access different real-time video statistical data and parameters. Our solution has been implemented only for Unix platform and Mozilla Firefox web browser for now, but in short time we

would like to target also Windows platform and all web browsers as well. Each client running operating system of Unix type must have a compiled version of VLC Multimedia plug-in installed in Mozilla web browser. This is necessary to work properly. This is a little disadvantage to bother clients with installing our modified multimedia plug-in, but this step can be easily implemented by downloading a debian package from a website and installation of package is simple and familiar to Unix users. What is more, our modified plug-in supports all of the functionality that is provided by original VLC multimedia plug-in, so there is no problem with compatibility.

2) VLC Multimedia Plug-in implementation: The source code of VLC Multimedia plug-in is written in C and C++ programming language. The code documentation of VLC media player can be found at http://www.videolan.org/developers/vlc/doc/doxygen/html/. Unfortunately this documentation is incomplete and does not include VLC multimedia plug-in. This makes our work a little complicated. The files that contain source code of this plug-in are located in projects/mozilla folder in downloaded source code files hierarchy. We also modified the sources of VLC media player in file media_player.c in src/control folder and control.c in src/input folder. Variables that store raw statistical data that are processed by input are in input_stats_t struct and the structure is shown in [Fig. 2]. These are only some statistical and information data about streamed video. Using Javascript API to exploit functionality of VLC Multimedia plug-in is documented at http://wiki.videolan.org/JavaScript_API.

Fig. 2.   input_stats_t Struct Reference

V. Conclusion

With the increasing trends in popularity of multimedia streaming services, combined voice, video and data services; video has significant impact on the whole new way how we work, live, learn and communicate. Another aspect that is constantly changing is the user's mobility. Therefore, the corresponding multimedia services have to be highly reliable to earn the trust of theirs customers. This can be only done by precise assessment of various aspects that have impact on resulted user satisfaction.

In order to achieve this goal we have to focus more on assessing of user experience. The proposed tool take advantage of high performance tool to gather real-time client-side streaming video statistical data on http streaming architecture. Data are retrieved and processed at client side, so the streaming server is not unnecessary stressed. This data are consequently sent to remote database where they are stored and linked to particular client. This helps to identify and solve problems with streaming service and improve quality of service. The tool is built on open source software, which is flexible, scalable and still improving. The tool can be used for testing and educational purposes. In future we have intention to extend it with interactive explanation of sequences within process of video quality evaluation.

References

[1] S. Winkler, *Digital Video Quality: Vision Models and Metrics.* London: John Wiley & Sons, 2005.
[2] D. Soldani, M. Li, and R. Cuny, *QoS and QoE Management in UMTS Cellular Systems.* London: John Wiley & Sons, 2006.
[3] F. Pereira, "Sensations, perceptions and emotions towards quality of experience evaluation for consumer electronics video adaptations," in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, January 2005.
[4] F. P. Speranza, F. Renaud, R. Caron, and M. Dupras, "Objective and subjective quality assessment with expert and non-expert viewers," in *Quality of Multimedia Experience (QoMEX)*, June 2010.
[5] A. Moorthy, K. Seshadrinathan, R. Soundararajan, and A. Bovik, "Wireless video quality assessment: A study of subjective scores and objective algorithms," in *Circuits and Systems for Video Technology*, April 2010.
[6] Huynh-Thu, Q. Garcia, M.-N. Speranza, F. Corriveau, and P. Raake, "Study of rating scales for subjective quality assessment of high-definition video," *Broadcasting*, vol. 99, 2010.
[7] U. Reiter and J. Korhonen, "Comparing apples and oranges: Subjective quality assessment of streamed video with different types of distortion," in *Quality of Multimedia Experience*, July 2009.

Systems based on 3D interfaces

František HROZEK

Department of Computers and Informatics, FEEI TU of Košice, Slovak Republic

frantisek.hrozek@tuke.sk

Abstract—**3D interfaces are becoming an everyday part of our lives. They are commonly used in medicine (surgeon training), manufacturing of various devices (prototypes printing), architecture (stereoscopic 3D visualization), entertainment (games controlled with motion) and many others. These 3D interfaces can be also combined together to create various systems with a new area of application. DCI FEEI TU of Košice is working on several such as systems and this paper aims on these systems.**

Keywords—**3D interfaces, 3D interfaces technologies, systems with 3D interfaces**

I. Introduction

3D interfaces are becoming an everyday part of our lives. Each one 3D interface has a wide range of application (3D displaying systems are used in visualization and presentation; position tracking systems are used for sensing devices position and for computer controlling; etc.). But combination of these 3D interfaces allows creation of systems which have a large new area of application.

DCI FEEI TU of Kosice is currently working on several systems that are based on combination of 3D interfaces. These 3D interfaces are: 3D scanner, data gloves, head mounted display, position tracking systems, 3D displaying systems, augmented reality, 3D printer and touch screen. These 3D interfaces are part of Center of Information and Communication Technologies for Knowledge Systems.

This paper aims on systems based on combination of 3D interfaces. Also aims on technologies of individual 3D interfaces located at DCI FEEI TU of Košice.

Paper is organized as follows. Section 2 and 3 is about 3D interfaces at DCI FEEI TU of Košice and about their technologies. In section 4 are mentioned systems that can be created from 3D interfaces located at DCI FEEI TU of Košice. Section 5 summarizes information presented in the paper.

II. 3D Interfaces and their Technologies

A. 3D Scanners

3D scanner is a device that analyzes a real-world object or environment to collect data on its shape and possibly its appearance (i.e. color).

There are several types of 3D scanners, which differ in the technology used for obtaining a data. They can be divided into two main categories: *contact* and *non-contact* scanners.

Contact scanners require a physical contact with the object being scanned. Although they are usually very precise, they are also much slower (order of 10^3 Hz) than non-contact scanners (order of 10^5 Hz).

Non-contact scanners use radiation to acquire required information about objects. They are of two basic types: *passive* and *active*. The main advantage of passive scanners is that they are cheap as they do not require so specialized hardware to operate. To scan objects, they only use existing radiation in its surroundings (usually visible light). In contrast, active scanners are equipped with their own radiation emitter (usually emitting laser light). While the latter are considerably more expensive, they are also much more accurate and able to scan over much bigger distances (up to few km) [1].

B. Data Gloves

Data glove is device, which serves for capturing mechanical and gestural information from the hand. Various technologies are used for capturing this information from the hand. These technologies can be divided into two categories: *determining the shape of the hand* and *position tracking*.

For determining the shape of the hand and fingers are used various types of bending sensors. The three most common types of bending sensors are conductive ink-based, fiber-optic based and conductive fabric/thread/polymer-based [2].

For sensing the hand position and rotation in the 3D space is used position tracking. More details about position tracking in the following section (II.D).

C. Head Mounted Displays

A head mounted display (HMD) is a display device, worn on the head (or as part of a helmet), that has small display optic in front of one (monocular HMD) or each eye (binocular HMD). HMD are used in virtual reality and augmented reality applications.

Another classification of HMDs is based on how the user sees the real world. This classification divides HMDs into two categories: *immersive* and *see-through*.

Immersive HMDs block a user's real world view to present him with a view that is under the full control of computers and to make him believe that he is part of the virtual environment.

See-through HMDs optically superimpose virtual objects upon an existing scene to enhance rather than replace the real scene. In this case, additional information is provided, so the user's visual system suspends disbelief and he or she perceives the computer-generated virtual objects as part of the real environment.

See-through HMDs have two subcategories: *video see-through* (user sees displayed real world image with added virtual objects indirectly via camera – display system) and *optical see-through* (user sees displayed real world image with

added virtual objects directly through semi-transparent mirrors).

For sensing HMD position and orientation is used position tracking. More details about position tracking in next section.

D. Position Tracking

Position tracking serves for sensing user's position and rotation in 3D space and can be used directly as 3D interface (for example Microsoft Kinect [3]) or be part of some 3D interface (for example in data glove).

Position tracking devices are divided into these categories (depending on technology used for position and rotation tracking) [4]:

- Mechanical trackers
- Magnetic trackers
- Ultrasonic trackers
- Optical trackers
- Inertial trackers

Each one tracking technology has its advantages and disadvantages. For example, magnetic trackers do not suffer obscuration problems, but they are sensitive to environmental magnetic fields and ferromagnetic materials in the workspace.

Tracking devices can be also combined, for example tracking system IS-900 SimTracker [5] combine ultrasonic tracking with inertial tracking.

E. 3D Displaying systems

3D displaying systems allows presentation of information in three dimensions. There are several technologies for 3D displaying and each technology has its advantages and disadvantages.

Displays used in these systems can be divided into [6]:

- Holographic displays
- Volumetric displays
  - Swept-volume displays
  - Static-volume displays
- Stereoscopic displays
  - Passive
    - Anaglyph
    - Polarized stereo
    - Infitec (**In**terference **fi**lter **tec**hnology)
  - Active
  - Autostereoscopic
    - Parallax barrier
    - Lenticular lens

F. Augmented Reality

There are several definitions of augmented reality [1]. One was given by Ronald Azuma in 1997. Azuma's definition says that Augmented Reality: combines real and virtual, is interactive in real time and is registered in 3D.

Another definition is by Paul Milgram and Fumio Kishino: Milgram's Reality-Virtuality Continuum. Continuum is visualized as line that is between reality and virtuality (Fig. 2).
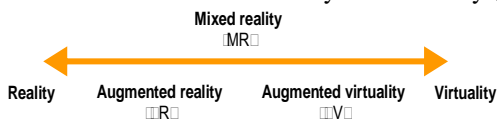

Fig. 2. Visualized Milgram's Reality-Virtuality Continuum

This continuum was extended by Mann into a two-dimensional plane of "Virtuality" and "Mediality" (Fig. 3). Taxonomy of Reality, Virtuality and Mediality.
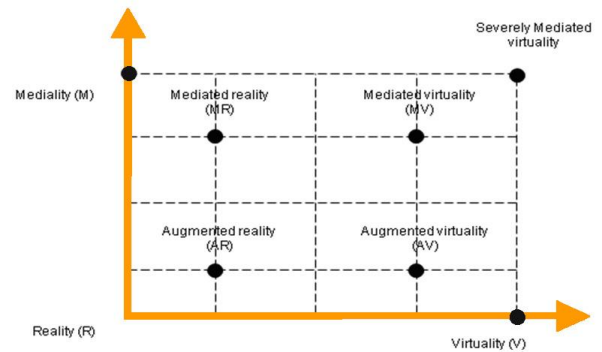

Fig. 3. Mann's Mediated Reality continuum

Based on how user sees augmented reality there can be two types of systems:

- Optical see-through systems
- Video see-through systems

According to method how virtual objects are aligned with real scene image there are two systems used:

- Marker systems
- Markerless systems

G. 3D Printing

3D printing is a form of additive manufacturing technology where a three dimensional object is created by laying down successive layers of material. 3D printers are generally faster, more affordable and easier to use than other manufacturing technologies (Stereolitography, Selective Laser Sintering and others). 3D printers offer developers the ability to print 3D models for visualization, testing or direct parts creation. Fig. 4 shows printed model of TUKE University library.
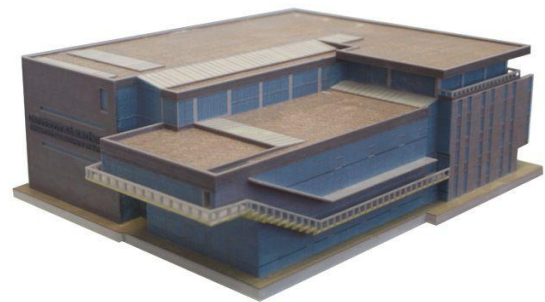

Fig. 4. Printed model of University library

H. Touch Screens

A touch screen is an electronic visual display that can detect the presence and location of touch within display area. Touch screen has two main attributes. First, it allows user to interact directly with what is displayed. Secondly, it allows controlling without requiring any intermediate device that would need to be held in the hand.

There are thee types of touch screens, which differ in technology used for sensing finger position [7]:

- resistive touch screens – use resistivity
- capacitive touch screens – use capacitance
- infrared touch screens – use infrared light

III. 3D INTERFACES AT DCI FEEI TU OF KOŠICE

Several 3D interfaces are located at DCI FEEI TU of Košice. These interfaces are part of Center of Information and Communication Technologies for Knowledge Systems.

3D interfaces at DCI:

- 3D scanner Leica ScanStation 2 (Fig. 5 left) [8]
- data glove DG 5 VHand 2.0 (Fig. 5 middle) [9]
- head mounted display nVisor ST60 (Fig. 5 right) [10]

Fig. 5. Leica ScanStation 2 (left), DG 5 VHand 2.0 (middle), nVisor ST60 (right)

- position tracking system Patriot – this system consists from three parts: control unit, source of magnetic field and sensor (Fig. 6) [11]

Fig. 6. Patriot tracking system parts (from left to right): control unit, magnetic field source, sensor

- position tracking system IS-900 SimTracker – this system consists from these parts: control unit, MicroTrax devices (head and hand tracker, wireless transmitter, wand – see Fig. 7) and SoniFrame [5]

Fig. 7. MicroTrax devices (from left to right) – wand, hand tracker, head tracker, wireless transmitter

- stereoscopic system that use INFITEC technology [12] – this system consists from: rendering cluster, projection screen, pair of projectors with special filters, glasses and visualization software (Fig. 8)

Fig. 8. Projection screen, glasses and projectors pair

- autostereoscopic 3D television Philips WOWvx (Fig. 9 left) [13]
- 3D printer ZPrinter 450 (Fig. 9 middle) [14]
- touch screen Microsoft Surface (Fig. 9 right) [15]

Fig. 9. Philips WoWvx (left), ZPrinter 450 (middle), Microsoft Surface (right)

IV. SYSTEMS BASED ON 3D INTERFACES

Each one of these 3D interfaces (mentioned in section III) has a wide range of application (3D displaying systems are used in visualization and presentation; position tracking systems are used for sensing position of devices and for computer controlling; etc.). But combination of these 3D interfaces allows creation of systems which have a large new area of application.

DCI FEEI TU of Košice is currently working on several systems that are based on combination of 3D interfaces mentioned in section 3. These systems are:

- System for sign language recognition
- System for augmented reality
- System for 3D visualization
- System for 3D printing
- System for interaction with virtual scene

A. Sign language recognition

System consists from pair of data gloves (VHand 2.0) and computer with application which transforms data obtained from data gloves (sign language signs) into spoken words (Fig. 10).

Usage: sign language recognition

Computer + application   Pair of VHand 2.0 gloves

Fig. 10. Scheme of sign language recognition system

B. Augmented Reality

System consists from Microsoft Surface (MS) and data helmet nVisorST60 (Fig. 11).

Microsoft Surface   nVisor ST60

Fig. 11 Scheme of augmented reality system

MS allows effective manipulation with 2D content, but with data helmet this 2D content can be extended to 3D. For example MS can display 2D markers. These markers are recognized with augmented reality system and through data helmet user see 3D objects placed over 2D markers. Touch screen of MS also allows easy manipulation with these 3D objects through 2D markers.

Usage: robotics, architecture, presentations, design, etc.

C. 3D visualization

This system uses for 3D visualization autostereoscopic 3D television Philipse WOWvx. Stereoscopic system based on INFITEC technology can be also used, all depends on user's and visualization needs. For example WOWvx doesn't need glasses, but has smaller displaying area (comparing to stereoscopic system). On the other hand INFITEC stereoscopic system has large displaying area, but for watching are needed special glasses.

Another very important thing in visualization is 3D model creation. This model can be created with:

- 3D modeling applications
- 3D scanner (Leica ScanStation 2)
- combination: 3D scanner + 3D modeling applications

Especially combination of 3D scanner and 3D modeling application is very suitable for creation of 3D model which is based on real scene (buildings, city areas, etc.).

Usage: visualizations and presentations

D. 3D printing

This system is similar to 3D visualization system, because it also uses 3D scanner and 3D modeling applications for 3D model creation. Difference is in the output, this system uses 3D printer ZPrinter 450 as output interface.

Usage: visualization of buildings (Fig. 12) and prototypes, devices and machines parts printing, etc.

Fig. 12. St. Michael Chapel printed model

E. Interaction with virtual scene

This system consists from these parts (Fig. 13):

- tracking system IS-900 SimTracker – serves for user tracking and navigation in virtual scene
- data gloves VHand 2.0 – serves for interaction (grabbing, moving) with virtual scene
- stereoscopic system that use INFITEC technology – serves for 3D displaying of virtual scene

Virtual scene for this system can be created by 3D modeling application and 3D scanner.

Usage: interactive presentations and visualizations

IS-900 SimTracker + Pair of VHand 2.0 gloves + Stereoscopic system with INFITEC technology

Fig. 13. Scheme of system for interaction with virtual scene

V. CONCLUSION

3D interfaces bring a lot of benefits to our everyday life. Several 3D interfaces and their technologies were presented in section 2 and 3.

These interfaces can be used individually or can be combined together. These combinations of 3D interfaces allow creation of systems which have new area of application. DCI FEEI TU of Košice is currently working on several such as systems. These systems and their usage were presented in section 4.

Future work will be focused on:

- create augmented reality system that use GPS and can be used outdoor
- create system for scanning large geographical areas
- create systems that use other 3D interfaces (3D cameras, locomotion interfaces) and implement these 3D interfaces to existing systems

REFERENCES

[1] Hrozek, F.; Sobota, B.; Janošo, R.: Visualization with 3D Interfaces, International Scientific Conference on Computer Science and Engineering, CSE 2010, Proceedings of, vol. 1, no.1, pp. 328-335, September 20-22, 2010, ISBN 978-80-8086-164-3
[2] Hrozek, F.: Data glove as input device for computer, Journal of Information, Control and Management Systems, vol.8, no.4, pp. 329-334, 2010, ISSN 1336-1716
[3] Microsoft Kincect homepage, http://www.xbox.com/en-US/Kinect/
[4] Durlach, I. N.; Mavor, S. A.: Virtual Reality: Scientific and Technological Challenges, Washington, D.C.: NATIONAL ACADEMY PRESS, 1995, 556 p., ISBN 0-309-05135-5
[5] Intersence IS-900 SimTracker homepage, http://www.intersense.com/pages/20/14
[6] Xu, S.; Manders, C.M.; Odelia, T.Y.; Song, P.: 3D display for a classroom, Educational and Information Technology (ICEIT), 2010 International Conference on , vol.2, pp.V2-316-V2-320, 17-19 September 2010
[7] Hoye, T.; Kozak, J.: Touch Screens: A Pressing Technology, Tenth Annual Freshman Conference, University of Pittsburg, April 2010
[8] Leica ScanStation 2 homepage, http://hds.leica-geosystems.com/en/Leica-ScanStation-2_62189.htm
[9] DG5-VHand 2.0 homepage, http://www.dg-tech.it/vhand/eng/index.html
[10] NVIS nVisor ST60 homepage, http://www.nvisinc.com/product2009.php?id=5
[11] Polhemus Patriot homepage, http://polhemus.com/?page=Motion_Patriot
[12] INFITEC homepage, http://www.infitec.de/index2.php
[13] Philipse WOWvx homepage, http://www.business-sites.philips.com/3dsolutions/home/index.page
[14] ZCorp ZPrinter 450 homepage, http://zcorp.com/en/Products/3D-Printers/ZPrinter-450/spage.aspx
[15] Microsoft Surface homepage, http://www.microsoft.com/surface/

System Aspect Mining from Java Library

[1]*Michal VAGAC*

[1]Deptartment of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]michal.vagac@gmail.com

Abstract—**Program comprehension is important part of software maintenance. Before applying any kind of change, it is essential to correctly understand current state of an application. In most cases, a change is applied in source code what implies that source code must be available and an affected application must be shut down and after applying change restarted.**

This paper presents innovative approach to simplify Java program comprehension in order to improve maintenance. By analyzing the way of using known Java class libraries it is possible to build a model of usage of these libraries. It is possible to build this model on higher level of abstraction than implementation level.

Keywords—**Aspect-oriented programming, java class library, metalevel architecture, program comprehension, reverse engineering**

I. INTRODUCTION

Question of applying software change is important field of research. None program is perfect and sooner or later it must be changed for different reasons. The most common ones are bug fixing and adding a new functionality. Among the other reasons for change there are different kinds of improvements of inner system structure, changes in original requirements demanded from the system, or different constraints during initial system development [1], [2]. Software system maintenance and evolution consumes up to 80 percent of its lifetime [3]. Software evolution represents a need for continuing maintenance and development of software used in real world applications or to solve problems in real world domain [4]. It is more common to change existing system than to create a new one.

Software maintenance is a difficult and complex process. Before applying a change, it is important to correctly understand affected source code of the system. Without fulfilling this condition it is possible to break functionality of the system. To change the program, a programmer must read it and understand it. Improvements in ways of program comprehension affect also program maintenance. Computer program is subject not only to computer, but also to its developer – a programmer.

There are several research fields treating a program comprehension. Among others two very important are reverse engineering and techniques of code analysis. Reverse engineering is a process of analyzing a subject system to create representations of the system at a higher level of abstraction [5], [6]. This process can be seen as coming back in development life-cycle. With a help of reverse engineering it is possible to get from implementation phase back to analysis phase. A knowledge acquired during this process can be used on maintenance of the system or on development of new one. The reverse engineering itself is just process of examination – the subject system is not changed. One of results of software engineering are different code analysis techniques. Among others there is a program slicing [7], [8], [9].

II. ASPECT-ORIENTED PROGRAMMING

Aspect-oriented programming allows modularization of crosscutting concerns. These concerns crosscut basic program functionality and are impossible to modularize using conventional ways (e.g. provided by object-oriented programming).

The term *aspect* describes concern which is impossible to describe by available language constructs (as classes or methods) used to modularize code. These concerns, named *crosscutting concerns*, are impossible to modularize from its essence, since they crosscut basic system functionality. This results in tangled code, which it is difficult to understand and reuse. Two concerns crosscut when they have to be composed differently, but same time they must be coordinated [10]. Aspect allows modularization of crosscutting concerns, which results in simpler code which it is easier to maintain.

Join points are points in which aspect crosscuts basic program. Composing defined aspects and affected program is task of aspect weaver. By defining a *pointcut*, it is possible to define whole group of joint points. *Advice* allows to define action executed at points defined by pointcut. Aspect represents modular unit consisting of pointcut and advice. The most common way of aspect implementation is weaving aspect code into program code.

III. A METHOD OF METAMODEL CREATION

An important aspect of developing complex software system is used programming language and related language libraries. The advantage of using such libraries is no need to develop common tasks all the time.

A program developed in object-oriented language is typically defined by group of classes and their instances - objects. When reading such a program, a programmer reads statements, control structures and used classes. If a programmer is not familiar with the used classes, it is required to get detailed information of all these classes. Another benefit of using classes from standard libraries beside reusing existing code is easier program comprehension.

In this section we propose a tool which – according to libraries used in the program – automatically creates a model of that part of program which uses these libraries. Since subject of the tool is a program, it is very convenient to use metalevel architecture. Base level will be presented by the program itself, while metalevel will contain created metamodel.

To get information about used classes, a process of examining base system is needed. Aspect-oriented programming together with better modularization allowed also extending
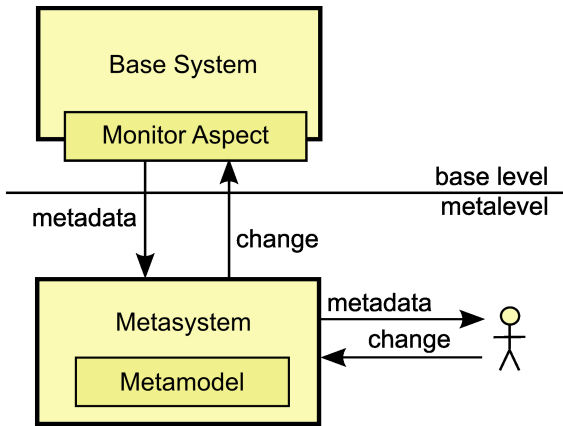
Fig. 1. Metamodel creation and propagation of change using aspect-oriented programming

existing code with a new functionality [11]. It is even possible to add a new functionality without access to source code. These properties of aspect-oriented programming were crucial for choosing this paradigm as a way of getting information about the base software system. With a help of aspect-oriented programming, the base system is extended with a new code, which has access to internal structures of the base system. This way it is possible to get information about the base system (using of known library classes) and on the basis of these information it is possible to build metamodel. Figure 1 represents basic architecture of described metasystem.

Metamodel represents known state of way of using specified library in the base system. As mentioned above, a crucial property is that because of known way of implementation of this library it is possible to build metamodel on higher level of abstraction. The advantage is that metamodel presented to user contains only information related to specified problem and no details of implementation.

IV. EXPERIMENTAL TOOL

Based on the method presented in previous chapter we developed two experimental applications. First one presents base level – an application modeling real world situation. This application implements simple address book. Second application is experimental tool placed in metalevel. This tool monitors base level application behavior.

Both applications are developed in Java programing language. Main reasons for choosing this language were its widespread use and rich standard library. The language in its base version contains library (named Java Class Library) with big amount of classes which help to solve wide range of common problems.

A. Base level application

Base level application presents simple situation – persons administration. The application is controlled by graphical user interface (Fig. 2). A user is able to search for person, create new person or update or remove existing person. Each person is identified by personal number, name, surname and list of phone numbers One important function to mention is possibility of import date from external system. External system is represented by standalone server application. After invoking import function, the base level application connects via socket to external system and imports provided data.
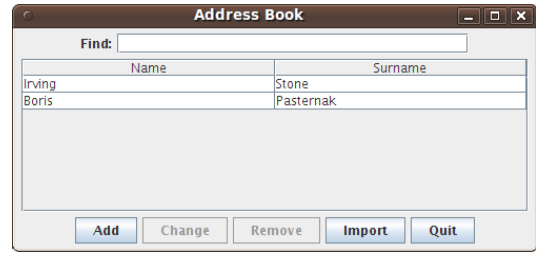
Fig. 2. Main window of base level application

TABLE I
BASIC METHODS FOR READING FORMATTED INPUT STREAM

Method name	Description
hasNext	checks if there is any symbol on input stream
next	reads next symbol
nextBoolean	reads next symbol as a boolean
nextByte	reads next symbol as a byte
nextDouble	reads next symbol as a double
nextFloat	reads next symbol as a float
nextInt	reads next symbol as a int
nextLong	reads next symbol as a long
skip	skips next symbol
useDelimiter	defines symbol delimiter

B. Data import process

As mentioned in previous section, base level application has ability of data import from external system. The application connects to external system via network socket. After successfully reading the data, these are parsed and then converted to application internal objects. Transfered data are represented as text. Each line of the text represents one person. Different values for person (as name, surname, etc.) are separated by space character. This format of data is strictly defined – changes in this format in any side of communication will break compatibility of peers.

This experiment deals with change of data format in external system side of communication. If external system adds new element (or remove existing), all import will become unusable – because of different data format sent than expected. The task is to cope with this situation.

C. Data import implementation

Parsing data read from external system is done with *java.util.Scanner* class. This class reads data from specified input stream line by line. Each line is split by defined delimiter to number of fields. Each field is automatically converted to specified data type. Table I describes base methods of this class.

D. Data format metamodel

It is possible to model described situation with table where each column represents data field (e.g. name, surname) and each row represents one record (a person). From metamodel point of view the rows are not important. Since the task is to enable data format modifications, it is needed to focus on the columns – used data types (Fig. 3). Changing data format means removing columns, adding new columns or change type of existing column.
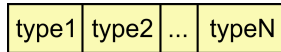
Fig. 3. Data format metamodel

E. Metalevel construction

To create a metamodel of data format it is needed to track down all invocations of methods of *java.util.Scanner* instance. All invocations of methods which name starts with *next* (we will refer to these methods as *next** methods) define expected format of parsed data. End of line of table is represented by *hasNext* method. Since only one instance of *java.util.Scanner* class represents all work with input stream, it is easily possible to find context between different method calls.

From aspect-oriented programming point of view there must be defined pointcut on creating new instance of *java.util.Scanner* class. At this moment new context and new metamodel are created. Other pointcuts are defined on invocations of *next** and *hasNext* methods. On execution of these methods current context is received from *java.util.Scanner* instance. There is stored current metamodel, which is built according to used *next** methods.

From change point of view there are three different possibilities:

- new type in data format
- missing existing type in data format
- changed existing type in data format

When there is a new type introduced in data format, the problem would arise because number of expected symbols and number of received symbols would differ. Let's take (T_1, T_2) as original data format. After introducing T_n data format changes to (T_1, T_n, T_2). This would lead to crash in base level application in a moment where the application would be expecting end of line but still be receiving another symbol (T_2). Other problem could arise with data type mismatch (receiving T_n instead of T_2). This problem can be solved by automatic insertion of proper *next** method invocation on place where new data type appeared.

Opposite situation happens when removing existing data type. After removing T_1 from previous example, one problem would be again data type mismatch (receiving T_2 instead of T_1) and another will be missing data (receiving end of line instead of T_2). It is possible to solve this situation by ignoring specified *next** method call. After invocation of the method on same position as removed data type, with help of aspect oriented programming this method will be ignored (not invoked). As a result the parser will remain on same position and following *next** invocation will read correct symbol.

Last situation – of changing existing data type – can be solved in similar way. After invoking *next** method which is on the same position as changed data type, it is needed to ignore this invocation and invoke different type of *next** method instead.

F. Tool implementation

Experimental tool was implemented according to presented assumptions. Base level application remains completely independent from the tool. All the connection between metalevel and baselevel is solely realized with help of AspectJ aspects. The tool allows to track down communication data format. Resulting metamodel is presented as list of used data types. It
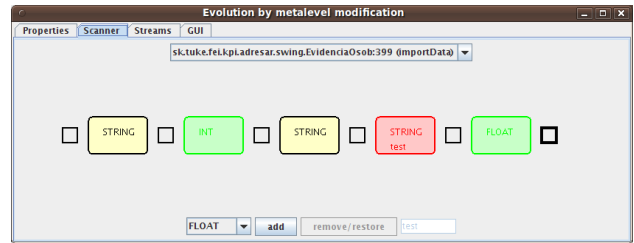

Fig. 4. Tool for changing communication data format

is possible to add new fields or to remove existing ones (Fig. 4).

V. CONCLUSION

With the experiment tool is is possible to automatically create metamodel of network communication format between the base level application and external system. This metamodel is created at higher level of abstraction than the base level implementation, so it simplifies the program comprehension. By metamodel changes it is possible to apply change to the base level application without a need to understand all details at implementation level.

The presented tool allows applying changes during system runtime – there is no need to stop the base level application. All changes are implemented only by changes in application object model. It is even possible to apply changes on application without need of source codes.

The implemented tool depends on Java class library. Therefore it is possible to use same tool on any Java application using same version of the standard library.

REFERENCES

[1] M. M. Lehman, "Laws of software evolution revisted," in *In: Proceedings of the 5th European Workshop on Software Process Technology, EWSPT'96*. United Kingdom: Springer-Verlag London, October 1996, pp. 108–124.

[2] M. M. Lehman, J. F. Ramil, P. D. Wernick, D. E. Perry, and W. M. Turski, "Metrics and laws of software evolution - the nineties view," in *In: Proceedings of the 4th InternationalSymposium on Software Metrics, METRICS'97*. Washington, DC, USA: IEEE Computer Society Press, 1997, pp. 20–32.

[3] M. M. Lehman, J. F. Ramil, and G. Kahen, "A paradigm for the behavioural modelling of software processes using system dynamics," Department of Computing, Imperial College, London, United Kingdom, Tech. Rep., September 2001.

[4] M. M. Lehman and J. F. Ramil, "An approach to a theory of software evolution," in *In: Proceedings 4th International Workshop on Principles of Software Evolution, IWPSE'01*. NY, USA: ACM Press, 2001, pp. 70–74.

[5] E. J. Chikofsky and J. H. C. II, "Reverse engineering and design recovery: A taxonomy," in *Journal IEEE Software*. Los Alamitos, CA, USA: IEEE Computer Society Press, January 1990, pp. 13–17.

[6] M. L. Nelson, "A survey of reverse engineering and program comprehension," in *In ODU CS 551 - Software Engineering Survey*, April 1996, p. 2.

[7] M. Weiser, "Program slicing," *ICSE '81 Proceedings of the 5th international conference on Software engineering*, 1981.

[8] F. Tip, "A survey of program slicing techniques," *JOURNAL OF PROGRAMMING LANGUAGES*, vol. 3, pp. 121–189, 1995.

[9] B. Xu, J. Qian, X. Zhang, Z. Wu, and L. Chen, "A Brief Survey Of Program Slicing," *ACM SIGSOFT Software Engineering Notes*, vol. 30, March 2005.

[10] G. Kiczales, J. Lamping, A. Mendhekar, C. Maeda, C. Lopes, J. marc Loingtier, and J. Irwin, "Aspect-oriented programming," in *In: Proceedings of the 11th European Conference on Object-Oriented Programming, ECOOP'1997*, vol. 1241. Jyvskyl, Finland: Springer-Verlag Lecture Notes in Computer Science, June 1997, pp. 220–242.

[11] M. Oriol, W. Cazzola, S. Chiba, and G. Saake, "Getting Farther on Software Evolution via AOP and Reflection," *Report on the 5th RAM-SE Workshop at ECOOP'2008*, 2009.

The assembly line model at Department of Cybernetics and Artificial Intelligence

1 *Matej ČOPÍK,* 2 *Ján ILKOVIČ,* 3*Tomáš KAROĽ*

1,2,3Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

1matej.copik@tuke.sk, 2jan.ilkovic@tuke.sk, 3tomas.karol@tuke.sk

Abstract— **The main objective of this article is a basic description of the assembly line located at the Department of Cybernetics and Artificial Intelligence. It is a practical educational model designed to improve the quality of the teaching process and also for research purposes. The description is focused on essential division of assembly line to functional blocks - posts. The assembly line has its global aim, and every post has its well defined role, so that the resulting aim is achieved.**

Keywords—**Assembly line, technological communication networks, PLC (Programmable Logic Controller), educational model.**

I. INTRODUCTION

In the framework of projects supported by EU Structural Funds are at the Technical University in Košice built research centers of excellence [4]. Under the leadership of Doc. Ing. Ján Jadlovský CSc. the Department of Cybernetics and Artificial Intelligence has completed assembly line named Flexible Assembly Company (FAC). FAC model also with the another assembly line [3] are located in laboratory V147 at Vysokoškolská 4. Its location in this room is also shown in Fig.1.
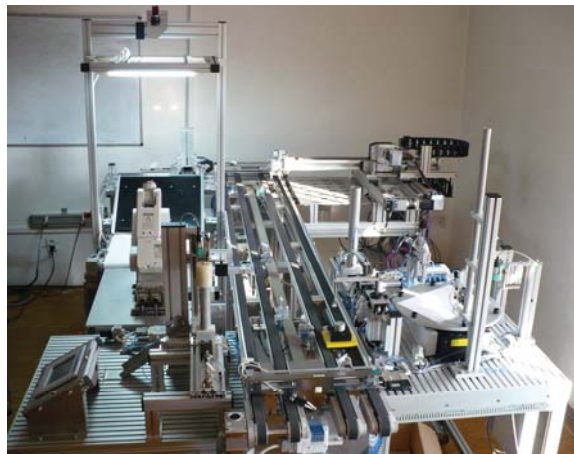
Fig. 1. The assembly line located in the premises of the department.

FAC model will serve as support for research and diagnosis of of industrial communication networks, the modelling of production lines, but the main task of this model is to improve the quality of the teaching process. Therefore, the model was specifically designed and this fact explains the details and selected technologies, which in practice would be better to choose alternatively to optimize the reliability and simplify the assembly process. The purpose of the model, however, was not to optimize proposal and a subsequent programming of the model, but summarize in one place as much possible technological elements as it possible. At this case students have the opportunity to become acquainted with multiple technologies and solutions, which are used in practice, and all in the one place. The FAP model obviously does not contain all the technologies used in practice, but the designers decided to use just those which they themselves frequently use in practice, and with which would be able to meet also graduates of our department in practice in the future. Some information about the technologies and design of this model are described in the following parts of the article.

II. ASSEMBLY LINE OUTPUT

The aim of the assembly process FAC is to assemble a product that is shown in Fig.2 (B). The resultant product consists of four intermediate products (Fig.2 (A)) that are assembled together by the assembly line in such way that product on pallette comes through the process posts and the intermediate products are progressively mounted together until a resultant product is completed. After its completion on output of assembly line is saved to serviser and awaiting to order allocation. The entire assembly process can be divided into five positions – posts (Fig.3). The entire assembly operations are divided between these five posts till to store the finished product warehouse. Each post has a well defined action, and these are directly linked to each other, i.e. that each post is waiting to close on the previous post, except the post number 1 (Fig.4), where the entire assembly operation
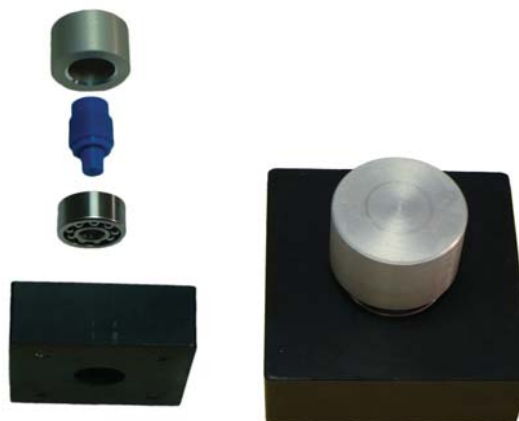

Fig. 2. Individual parts of the product in the order they are assembled (A). The resulting product (B).

begins. In the fully operation of assembly line it should not happen, when would any post was in work and others should be idle because the design and hardware assembly line enables production of multiple products simultaneously. Matter of course is the synchronization of individual activities so that there is a collision in the production process. Consecutive posts have the ability to arrest the intermediate product on belt conveyor until then, when the next post released, although the previous post has already finish its activity. Assembly line work will be cyclically repeated, up to the fulfilment final product store. Products are transferred between the various posts by using a special palette (in the middle part of Fig.1 we can see palette with already finished products situated on the belt conveyor). Palette itself is therefore not part of the finished product and hence will not be placed in serviser. All pallets are situated on a belt conveyor and either they are waiting in the queue at the beginning of the assembly process, or move product between the posts assembly line.
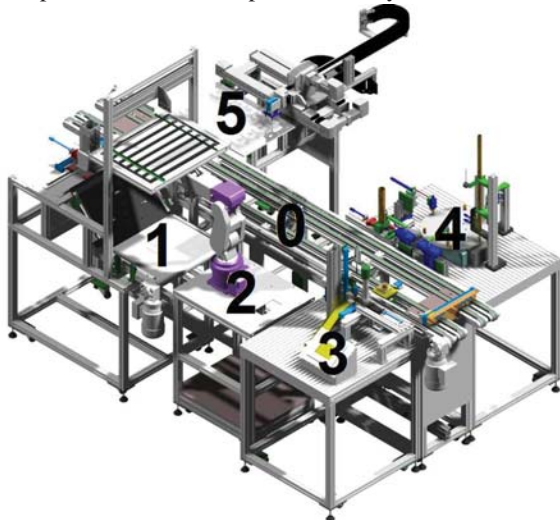
Fig. 3. Graphic design of the assembly line with the numbered posts.

III. DESCRIBED POST FUNCTIONS OF THE FLEXIBLE ASSEMBLY COMPANY

A. Post 0 – Belt conveyors

Belt conveyors are situated in the middle of the assembly line, as can be seen in Fig.3 (also in the Fig.1) and serve to transfer the product at each assembly line posts. Belt conveyors are powered by three-phase motors with a gearbox to drive the belt conveyors at a constant speed and constant direction. The conveyors are triggered in software and should move during the whole operation without stopping the assembly line. Stopping the pallets with products in the required position is indicated by the presence sensors and is done by using of pneumatic up-tongues that stop the palette at the required position on the belt conveyor, although it still runs. In this way is resolved the independence of assembly tasks (post) among each other, although they are located on the same belt conveyor (on the belt conveyor can move independently at the same time multiple pallets). If there is palette with products at the end of belt conveyor is necessary to move it to another belt conveyor. And because the belt conveyors are moving against each other, so the begin of the second belt conveyor is now near the end of the first one. At

both ends of these conveyors is therefore simple pneumatic device, which ensures pallets movement from the end of the first, the belt on beginning of second one.

B. Post 1 – Ejecting the blocks and camera system
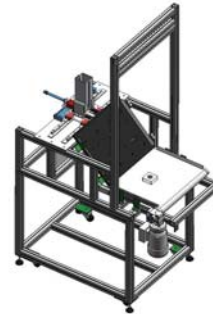
Fig. 4. Graphic design of the Post 1 – Ejecting the blocks and camera system.

On the Post 1 (Fig.4) starts the assembly process. This post contains a serviser for black plastic blocks into which will be inserted other parts of the product (bearing, plastic shaft and metal cap). Blocks are pneumatically removed from a special serviser on sloping surface that contains pins, which has randomly rotated the blocks. Subsequently, randomly rotated block gets on a wide white belt conveyor located behind mentioned sloping surface. Above the belt conveyor is color industrial camera with artificial lighting, which is connected to desktop computers, whose task is to process data from the camera and on the basis of these data to calculate the exact position and rotation of the block on a belt conveyor. Data from the camera system are forwarded to the central control computer PLC (Programmable Logic Controller), which uses the data for coordinate movements of industrial robot arm. To facilitate the image recognition has the belt conveyor a white color and it is powered by three-phase motor with gearbox, whose speed is controlled from a PLC via a frequency converter.
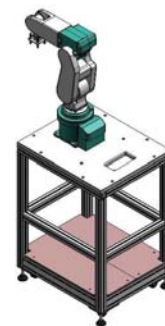
C. Post 2 – Robotic arm

Fig. 5. Graphic design of the Post 2 – Industrial robotic arm Mitsubishi RV-2SDB with the rack for placing of its control system.

From the first post and from the belt conveyor is the block grabbed by the the arm of robot (Fig.5) and thanks to the information from the camera system about the position and the rotation of blocks is the arm able to grab a block from anywhere on the conveyor and move it to an empty palette placed on a belt conveyor (Post 0). It is a dimensionally small industrial robot Mitsubishi RV-2SDB that has on the end of the arm positioned pneumatic suction, by using which is able to manipulate with the blocks. The arm design gives it a very

good movement abilities, by using which it can also move in a limited workspace. [2] Kinematic structure of the robot arm has six degrees of freedom powered by six servomotors, which give it very good dynamics and precision of movement. The arm is attached on a special control computer supplied by the manufacturer, which is an essential part of an industrial robot and directly manages and monitors the activities of arm. This control computer is physically located below the arm worktop and directly supplies the arm actuators. Mitsubishi robot by using its control computer can communicate with other control systems or devices the surrounding technological communication network. Even in the FAC application is robot connected to the control PLC by technological Profibus network, via which it receives control commands and information about the location and rotation of the product on a belt conveyor (Post 1) from the master PLC, which led it operates.

D. Post 3 – Inserting bearings and PanelView
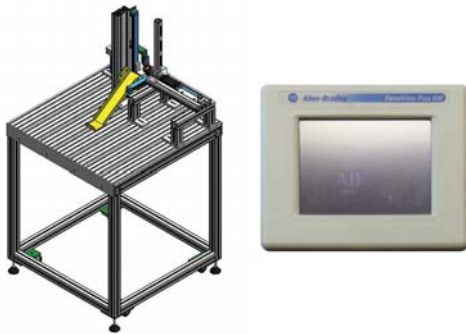
Fig. 6. Graphic design of the Post 3 – Inserting bearings rack (A). PanelView Plus 600 (B).

After the previous activities of the posts the block is loaded on a pallet belt conveyor and in it is necessary to insert the bearing. Bearing, however, must have the correct height. Maximum outer dimension of the bearing is limited by serviser size. So bearing is necessary to pick from a serviser, to measure and insert it into the block. And these activities are carried out in the post 3 (Fig.6). Bearings are situated in the serviser similar to that on the Post 1, but this is in the shape of a hollow circular cylinder. At the bottom part of serviser are by the pneumatic piston removed new bearings. In the serviser of our application are bearing of two dimensions. Therefore, after ejecting the bearing from the serviser is always measured his height. If the dimension of the bearing is unsuitable then it is discarded and from the serviser is ejected the new bearing. This operation is repeated until from the serviser is obtained bearing with the correct dimensions and then by using a rotary pneumatic manipulator is inserted into a prepared block, which is already waiting on the conveyor. When the correct bearing is embedded into the block, then the palette with the product proceeds by the conveyor to the next post.

Post 3 include also a small touch screen PanelView Plus 600, which is connected to the controller PLC.

E. Post 4 – Inserting plastic shaft and metal hats.

Final assembly work is carried out at the Post 4 (Fig.7). Into the imbedded bearings in the plastic block is still imperative to insert a plastic shafts, on which is ultimately put metal cap. On Fig. 2 (A) are both parts are displayed in the upper part of the picture.

This post consists of 6-position rotary table - rotary conveyor, which is the central facility of the post. At each of its six positions can either insert plastic shaft or metal cap. Rotary conveyor thus serves the fast and accurate transfer of hats and shafts between the devices, which are located in this post. Under the title devices, it is necessary to understand the sensors, servisers and pneumatic handlers, whose job is to choose caps and shafts with the right parameters and then use those to complete the final product.
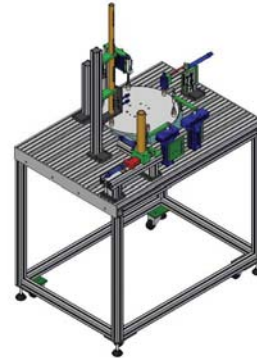

Fig. 7. Graphic design of the Post 4 – Inserting plastic shaft and metal hats.

Caps and shafts have its own serviser above the rotary table, and at each one there are two types of stored products. Operator must manually fills the serviser. From the serviser to the rotary table positions are these semi-product situated by pneumatic devices. Plastic shaft dimensions are the same, but are distinguished by the colour to green and blue. The application is operator fault-tolerant, because he could insert shaft into the reservoir rotated incorrectly. Metal caps are also dimensionally identical and also are together in the same serviser, but they differ in material from which they are made (metal and aluminium). Each tray is located above the one of the positions of the rotary conveyor, what makes together two used positions. Over the other two positions of the rotary conveyor are situated special sensors, videlicet sensor colour and sensor for measuring the height of product. Using the colour sensor, we find the colour of shafts and the other one sensor is used to determine whether the shaft is rotated in the correctly position, i.e. whether the operator inserted shafts into the serviser correctly. In the case of incorrect product embedding to the rotary conveyor, the product is discarded from the used positions using a pneumatic manipulator, which is located above another position of the rotary conveyor. The last available position uses a simple pneumatic manipulator with pneumatic suction providing moving shaft and hats from form the rotary conveyor to the waiting pallet on a belt conveyor (post 0). After the inserting of both components in the correct order on the block with bearing the product is finished and proceed to the final post 5.

F. Post 5 – Warehouse

Palette with the finished products stops at the last post and triaxial manipulator moves product from the pallet to its position in the warehouse (post 5 in the Fig.8). The triaxial manipulator is powered in two axes by stepper motors and the third axis is controlled pneumatically. The manipulator handles the products using suction cups. This manipulator was

selected after the positive experience in its use in a similar project before [3]. Warehouse is implemented as a metal table, where milled positions for products are marked with letters of the alphabet from A to Y. Together there are 24 positions. Final position for each product will be stored in the information system of the FAC.
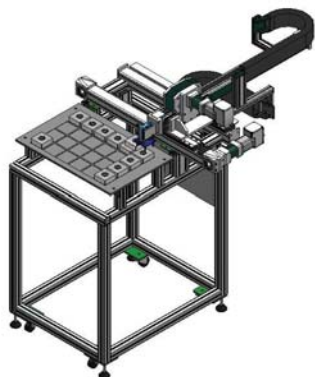

Fig. 8. Graphic design of the Post 5 – Warehouse of the resulting products.

Delivering finished product from warehouse carries line operator by hands, despite the fact that triaxial manipulator can moved it too, but only on to the belt conveyor Post 0.

IV. NETWORK CONTROL SYSTEM
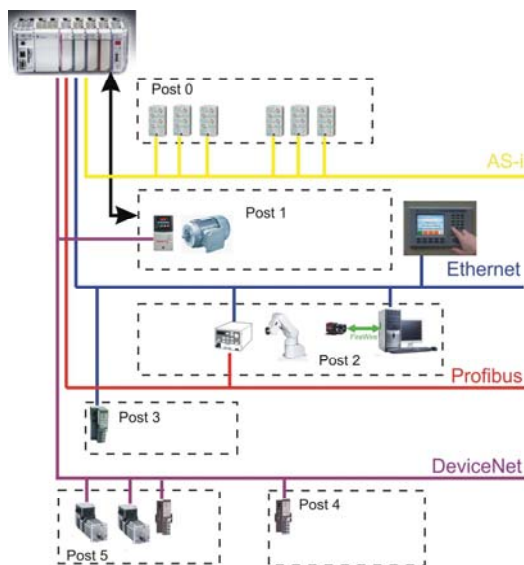
Fig. 9. FAC network control system.

The FAC is controlled by PLC which is equipped with network adapters. The most sensors, actuators and electrical devices of FAC control system are connected to the control unit via the technological networks. All technological posts except the posts two and one use remote inputs and outputs. There are together four types of technological networks used in this technological system (Fig.9): AS-i, Ethernet, Profibus, DeviceNet. This special system configuration of use the various types of technological networks is not common way of control system design, because is not easy to correctly configure the industrial network to work correctly. What more difficult task for a programmer is to configure up four networks in one control system. It is more easily to work only with direct PLC I/O. But it is caused by the use of this system. This system is mainly designed especially for industry network

research and also for learning at scholar process at scholastic environment.

V. CONCLUSION

The FAC model is now already a second assembly line at our department that was built by the EU financial support.

This model was to school supplied completely assembled but without the control programs, that are being developed at this time and there are also resolved basic model electrical and the construction problems and its debugging. In parallel are also developed other applications, namely: an application to image recognition, the control program of the industrial robot Mitsubishi, the visualization applications, the information system application and course it is solved communication and synchronization of all these applications.

Another important objective is to support research at our department and especially in the field of modelling and diagnosis of discrete systems. The authors of the article are also dealing with this issue in its doctoral dissertation [1].

For the FAC model design was used standard automation software and hardware technologies. Control system, sensors and actuators were also selected from the standard used industrial elements. Some of the technical and engineering solutions should be designed easier in practice, but this is special application.

All these and other solutions related to the fact that it is a educational model where a student has the opportunity to try out several technology equipment on a small model. As mentioned in the article introduction, the main objective is to improve the educational process and improve the adequacy skills and competencies of graduates of our department. Furthermore, solution is complicated so that all technologies have to be interconnected to one system and it have to be synchronized to work together to do the defined tasks. Using the more types of technologies in one system is for the programmer more complicated than using only one technology type. But if a student handles this task, thereafter in practice it will be more easily to him manage the complex tasks that it will be given to him. Based on the gained experience he will not be narrowly focused and so he will be more competitive at the labour market. And that's also an important task of the University.

REFERENCES

[1] J. Ilkovič, "Modeling and diagnosis of handling and robotic systems", The written part of dissertation examination, Košice, 2010 (in Slovak).
[2] Mitsubishi Elektronic – Mitsubishi industrial robot manual: RV-2SDB_2SQB_E.pdf: MELFA RV-2SDB/RV-2SQB Series, May 2010.
[3] ČOPÍK, Matej: Design and implementation of flexible manufacturing system management. Diploma work. Technical Univerzity in Košice, Faculty of Electrical Engineering and Informatics, 2010 (in Slovak).
[4] J. Jadlovský,S. Laciňák, J. Chovaňák, J. Ilkovič, "Proposal for Distributed Control System of Flexible Production Line ",Journal of Cybernetics and Informatics, Volume 11 published by Slovak Society for Cybernetics and Informatics, pp. 8 – 13, http://www.sski.sk/casopis/index.php, ISSN: 1336-4774, 2010.

The expressing of structures for computation in linear logic

Pavol MACKO

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

pavol.macko@tuke.sk

Abstract—**Algebras are mathematical structure that widely used in informatics. They are used for constructing basic structures used in computer programs. Coalgebras are dual structures to algebras and they are usually used for description of program behavior. In this contribution we present how to combine algebras and coalgebras and their properties for computing the factorial. By using the Curry-Howard correspondence and the proof-as-program paradigm we construct formulae of linear logic for proving the correctnes of the computation and we present the appropriate proof.**

Keywords—**anamorphism, catamorphism, hylomorphism, linear logic, Curry-Howard correspondence**

I. INTRODUCTION

Linear logic provides a logical perspective on computational issues such as control of resources and order of evaluation. The most important feature of linear logic is that formulas are considered as actions. While classical logic treats with the sentences that are always true or false, in linear logic it depends on an internal state of a dynamic system. Curry-Howard correspondence is a correspondence between logic and computing in informatics.

We have showed in [1] a new way of computing factorial based on hylomorphism by using coalgebras which is more effective. In this contribution I present correctness of this computing by Curry-Howard correspondence.

Usually the Curry-Howard correspondence [2] is used as follows: At first we make the proof of the problem and then we develop the program. But Curry-Howard correspondence is bijective, so it applies to both ways. We prove the correctness of the developed program. This opposite method is called reverse engineering. For checking the correctness of the program is the most important phase of transformation into logical formulas. Proof will be constructed in linear logic.

II. BASIC NOTIONS

A. Linear Logic

Linear logic has been introduced by Girard [3] as resource-sensitive refinement of classical logic. It is a strong system of logic and it has full features of first order and intuitionistic logic and additionally supports concepts of disposable resources and their consumptions. Linear logic provides a mechanisms to destroy and construct formulas in the process of proving, where formulas represent actions.
Linear logic provides a logical perspective on computational issues such as control of resources and order of evaluation. Using Curry-Howard correspondence, propositions of linear logic are interpreted as types. This paradigm has been a cornerstone of new approach concerning connections between intuitionistic logic, functional programming and category theory.

B. Curry-Howard correspondence

The Curry-Howard correspondence is the direct relationship between computer programs and proofs in constructive mathematics. Also known as Curry-Howard isomorphism and it forms the proofs-as-programs and formulae-as-types paradigm [2]. It is a generalization of a syntactical analogy between systems of formal logic and computational calculi. It was formulated by the American mathematician Haskell Curry and logician William Alvin Howard.

C. Algebras and Coalgebras

In our approach we treat with mathematical theory of programing. We use dual mathematical structures: algebras and coalgebras. Usually we work with them in the categories [4], [5]. There exists special algebra and coalgebra, namely an initial algebra and a terminal coalgebra.
It holds for initial algebra, that there exists the unique morphism from the initial algebra into any algebra. This morphism is called the *catamorphism*.
Dually, there exists terminal coalgebra, for which holds, that from any coalgebra exists unique morphism into terminal coalgebra, called *anamorphism* [6], [7].
Composition of those morphisms is a new morphism, which called the *hylomrphism*. We apply it in the effective computation of the factorial.

D. Hylomorphism

The hylomorphism recursion pattern was firstly defined in [9]. Given a functor F, it expresses the following recursive function of type $U \to A$ defined by fixpoint [11]

$$hylo(g, h)_F = \mu(\lambda f.g \circ Ff \circ h)$$

for the function $f : FA \to A$ and for the function

$g : U \to FU$.

Given an F-coalgebra $\varphi : U \to FU$ and an F-algebra $\alpha : FA \to A$, the hylomorphism denoted by $hylo(\alpha, \varphi)_F$ is the least arrow $f : U \to A$ that makes the following diagram at fig. 1 commute [10].
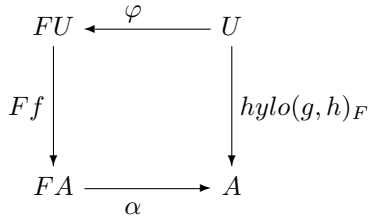
Fig. 1.   Diagram of hylomorphism

Moreover, the hylomorphism is a composition of an anamorphism with a catamorphism [10]:

$$hylo(\alpha, \varphi)_F = (cata\ \alpha)_F \circ (ana\ \varphi)_F.$$

III. THE TRANSFORMATION OF COMPUTATION TO THE LOGICAL PROOF

The program consists of data structures and algorithms. Data structures have their types [8]. Algorithm can be understood as a composition of operations, i.e. the relationship between typed data structures. Results of operations are obtained by evaluating typed terms. We can say that the program is modeled as the mathematical mapping correctly defined data domains.

The first step in the design solution is type theory that we will construct for a given problem. Types and operations with types we enclose into well-know from the universal algebra, many-typed signature.

$$\Sigma = (T, F)$$

The signature consists of a finite set of basic types T

$$T = \{int, intList\}$$

and a set of function symbols:

$F = \{ ==: intList, intList \rightarrow \Omega,$
$=: int, int \rightarrow \Omega,$
$join : int, intList \rightarrow intList,$
$* : int, int \rightarrow int,$
$pred : int \rightarrow int,$
$head : intList \rightarrow int,$
$tail : intList \rightarrow intList,$
$\}$

If the definition of types theories for the problem is fully, we are able to formulate terms and using by Curry-Howard correspondence we find type to each formula (prop-as-types). In our contribution we show a more effective method for computation of the factorial. For this we need terms, which represent catamorphism and anamorphism. Our function $fact(n)$ is based on hylomorphism, that is a composition of catamorphism and anamorphism. Function $fact$ consist of composition two functions, too. This functions are named as morphism, which describes their, namely: $cata$ and ana, resp.

1) The function $ana(x)$: By applying the anamorphism in the informatics we get a corecursive function that starts with a single input (here int) and returns it in a more complex output, for example a wide list ($intList$) and it is of type:

$$int \rightarrow intList.$$

Definiton of this function:

```
ana (x) =
if (x=0) then ana = emptyList
elseif (x=1) then ana = [1]
else ana = join(x,ana (pred x))
```

Term of this function has the following form:

$$n :\ int\ \vdash\ if\ (x = 0)\ then\ \varepsilon$$
$$elseif\ (x = 1)\ then\ [1]\ else\ join(x, ana(pred\ x))$$

Formula representing the function $ana(x)$ is:

$$(\varphi_1 \multimap \psi_1)\ \&\ (\varphi_2 \multimap \psi_2)\ \&\ ((\varphi_1^\perp \otimes \varphi_2^\perp) \multimap \psi_3)$$

2) The function $cata(l)$: By applying the catamorphism in the informatics, we get a recursive function that starts with a list (here $intList$) and returns it in a single output (here int) and it is of type:

$$intList \rightarrow int.$$

Definiton of this function:

```
cata (list) =
if list == emptyList then cata = 1
else cata = head(list) * cata (tail(list))
```

Term of this function has the following form:

$$l :\ intList\ \vdash\ if\ (list == \varepsilon)\ then\ 1$$
$$else\ head(list)\ *\ cata(tail(list))$$

Formula representing the function $cata(l)$ is:

$$(\theta \multimap \alpha)\ \&\ (\theta^\perp \multimap \beta)$$

3) The function $fact(n)$: The composition of functions ana a $cata$ creates a function to compute factorial $fact(n)$. The function generates a list of natural numbers by increments to number n, and simultaneously the list is eliminated by the multiplication operation between elements of the list. The function is of type:

$$int \rightarrow intList \rightarrow int.$$

Definiton of this function:

```
fact (n) = cata (ana (n)) =
if (ana(n) == emptyList) then fact = 1
else fact = n * cata (ana(pred n))
```

Term of this function has the following form:

$$n :\ int\ \vdash\ if\ (ana(n) == \varepsilon)\ then\ 1$$
$$else\ n\ *\ cata(ana(pred\ n))$$

Formula representing the function $fact(n)$ is:

$$((\varphi_1 \multimap \psi_1) \multimap \alpha)\ \&\ ((\varphi_2 \multimap \psi_2) \multimap \alpha)\ \&$$
$$\&\ (((\varphi_1^\perp \otimes \varphi_2^\perp) \multimap \psi_3) \multimap \beta)$$

Each atomic formula is constructed as a type. In formulas is seen the transition from type int to type $intlist$ and then back to type int.

IV. THE PROOF

The logical proof of the given formula in the previous section is at fig. 2.

Fig. 2.   Proof of formula expressing factorial computation

The tree of proof were constructed by using a natural deduction system and by using a logical rules of linear logic. In leaves of tree is captured computation of factorial, which it is true.

After the constructing of this tree, we can say, that our formula is proven, which means that our program is correct and does not need any verification.

V. Implementation in OCaml

A. The function $ana(n)$

This function is defined following: If the argument of the function ana is 0 then it returns an empty list. If the argument is 1, ana generates a list containing only 1 as item. Otherwise, ana generates a list with new element appended.
The implementation of the function $ana(n)$ is:

```
let rec ana n =
match n with
| 0 -> []
| 1 -> [1]
| n -> n :: ana (n-1);;
```

OCaml is a language, which after defining a function return the note about relationship between used types.

$$val\ ana : int \rightarrow intList$$

B. The function $cata(l)$

This function takes as its argument a list of factors of the type int and returns the result of multiplicative operations. The result of the function is an element of the type int, that is the result of multiplication of elements in the list.
The implementation of the function cata(l) is:

```
let rec cata list =
match list with
| [] -> 1
| head :: tail -> head * (cata tail);;
```

$$val\ cata : intList \rightarrow int$$

C. The function $fact(n)$

Composition of two function $cata \circ ana$ is written in programming language $OCaml$ as $cata(ana(n))$. The definition of this hylomorphism function $fact(n)$ is as follows:

```
let fact n =
cata (ana n);;
```

$$val\ fact : int \rightarrow intList \rightarrow int$$

Now we show example of the evaluation of function $fact$ with input value 4.
Execution of this function in OCaml:

```
# fact 4;;
- : int = 24
```

Illustration of this example step by step:

```
fact 4 =
     cata (ana 4) =
  4  cata (ana 3) =
 12  cata (ana 2) =
 24  cata (ana 1) =
 24 id =
 24
```

It is seen that method of programming by using the hylomorphism is more effective and it takes less steps.

VI. Conclusion

We presented in this contribution new way of the factorial computing. Our way is based on the algebraic and coalgebraic structures: the anamorphism, the catamorphism and the hylomorphism. These structures provide the computation which we proved whith Curry-Howard correspondence and we constructed the logical proof of the appropriate formulas in linear logic.
Our next goal is going to be the extension of this approach in other categorical structures based on recursive coalgebras and contructing of the appropriate description of the computation in linear logic.

REFERENCES

[1] P. Macko and V. Slodičák, "New approaches in functional programming using algebras and coalgebras," *The European Joint Conferences on Theory and Practice of Software (ETAPS) - Workshop on Generative Technologies*, 2011.

[2] M. Sørensen and P. Urzyczyn, "Lectures on the curry-howard isomorphism," 1998.

[3] J. Y. Girard, "Linear Logic," *Theoretical Computer Science*, vol. 50, pp. 1–102, 1987.

[4] M. Barr and C. Wells, "Category Theory for Computing Science," *PrenticeHall Int.*, 1990.

[5] V. Slodičák, "Some useful structures for categorical approach for program behavior," *Journal of Information and Organizational Sciences, University of Zagreb*, 2011 (accepted).

[6] J. Adámek, D. Lucke and S. Milius, "Recursive coalgebras of finitary functors," vol. 4, p. 447462, Oct. 2007.

[7] T. Uustalu, V. Capretta and V. Vene, "Recursive Coalgebras from Comonads," *Inf. Comput. 204*, vol. 4, p. 437-468, 206.

[8] J. Y. Girard, "Proofs and Types," *Cambridge University Press*, pp. 1–175, 2003.

[9] M. Fokkinga and E. Meijer, "Program calculation properties of continuous algebras," *Technical Report CS-R9104*, CWI, Amsterdam, January 1991.

[10] J. Kabanov and V. Vene, "Recursion Schemes for Dynamic Programming," *In T. Uustalu, ed., Proc. of 8th Int. Conf. on Mathematics of Program Construction, MPC06*, Tartu, Estonia, 2006.

[11] A. Cunha, "Point-free Programming with Hylomorphisms," *Universidade do Minho*, Braga, Portugal.

The majority method for diagnostics of the turbojet engine MPM 20

[1]*Tomáš KAROĽ*, [2]*Ján ILKOVIČ*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]tomas.karol@tuke.sk, [2]jan.ilkovic@tuke.sk

Abstract— **This article is focused on the diagnostics and backup. The individual sections are devoted to basic terms from these areas and the last deals with the design of the majority method for small turbojet engine (MPM 20).**

Keywords— **backup, diagnostics, small turbojet engine, MPM 20, majority method.**

I. INTRODUCTION

Nowadays, the terms like security and safety of systems are of increasing importance. This applies even more to the control of aviation engines and avionics systems, because when an error occurs, the machine can not be simply turn off. The control system has to ensure the implementation of the necessary functions of the aircraft to enable to maneuver and land safely. This is the area where the backup and diagnostics comes to the fore.

II. BACKUP

Its main role ([1], [5]) is fault tolerance, and this is achieved by using redundant resources, which we call backup. In implementing the backup, there are two basic relationships between the backed up and the backup element.

Division of backups:

Static - is working continuously with a system all the time. The input and output of system are permanently connected. Its main advantage is the ability to mask errors and impermeability of errors on the system´s output. On the other hand, its disadvantage is the increased energy consumption and low mean time between failure (because all elements are simultaneously worn out).

Dynamic - enters into system when the backup object gets to the error state. It is activated by using a switch. With dynamic reserve a longer mean time between failure and lower power consumption than with static reserve is achieved. However, the disadvantage is that during switching occurs a temporary loss of output signal. In this backup method three basic modes of operation are distinguished (it dependents in proportion to the size of performance in relation to backup and backed up device):

- **unloaded** - backup element - when it is not used - is completely disconnected from power (zero power),
- **lightweight** - backup element is turned on only at reduced power in the time, when the backed up

element is working,
- **loaded** - represents the opposite extreme to the unloaded backup (the backup works on 100% performance together with backed up element).

Hybrid - tries to connect the positives of the static and dynamic backup and eliminate their weaknesses.

III. DIAGNOSTICS

It is analysis of immediate state of the object ([1], [2], [3], [5]). It consists of two phases- fault detection and fault localization.

Diagnostic system (DS)

It is a system designed and focused to evaluate the state of the object. It consists of diagnostic tools (sensors, testers, diagnostic algorithms, modeling techniques, ...), diagnosed system and operator. Diagnostic tools can be implemented as a part of the object (internal) or separately (external resources). Division of diagnostic systems:

- ON-LINE DS - evaluates the continuous status of an object at work (Figure 1),
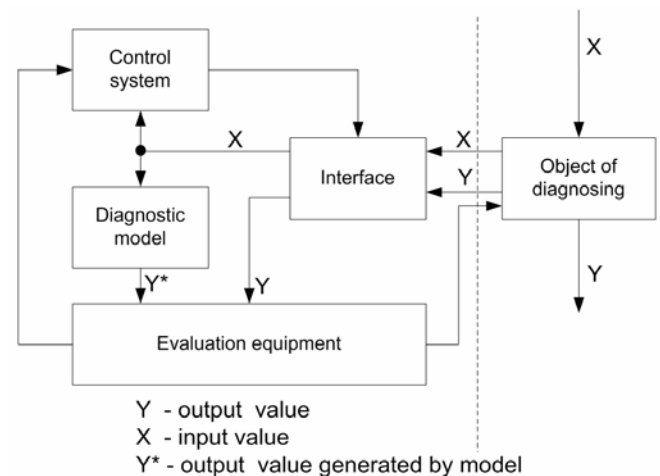

Fig. 1. ON-LINE diagnostic system

- OFF-LINE DS - diagnostics in this type of DS is most frequently realized by testing when the object of diagnosing is off (Figure 2).
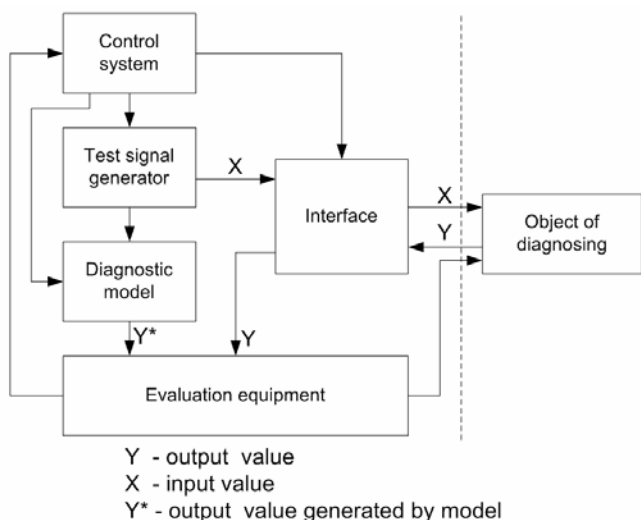
Fig. 2. OFF-LINE diagnostic system

Y - output value
X - input value
Y* - output value generated by model

Diagnostics has great importance for aircraft's engines and aircrafts, because if error occurred during the flight, this fault must be detected enough early to make the necessary actions for safe continue of the flight or landing. One of the primary methods used in practice (for diagnostics as well as for backup) is the majority method, which is based on the static backup (Figure 3) and has the ability to capture failures of input elements. Number of inputs fed into the evaluation facility must be odd to avoid the case where a number of false and correct signals are the same.
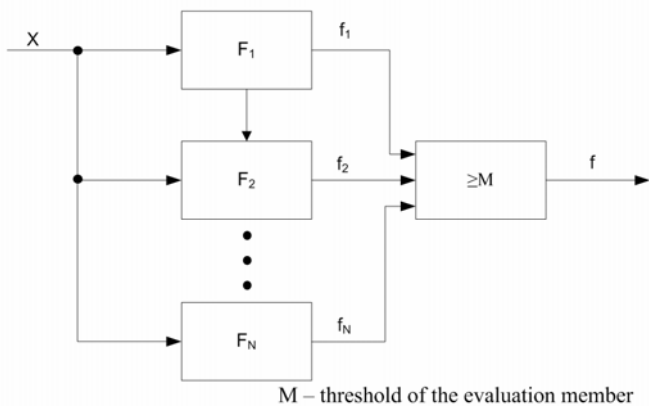
Fig. 3. General scheme of the NMR structure (n- modular redundant)

More detail description of proposal of the majority method see the next.

IV. THE MAJORITY METHOD FOR TURBOJET ENGINE MPM 20

It is online diagnostic system based ([4], [6], [7], [8], [9]) on static method of backup and used for checking the multiple backed up devices - to detect the failure of one of them. In this case we confine ourselves only to a system with three backups, but generally it can be n (odd number). In such a system we assume that the probability of failure is bigger for one of them, usually two work correctly (Figure 4). Measuring element, whose task is to detect the faulted element is called a quorum element.
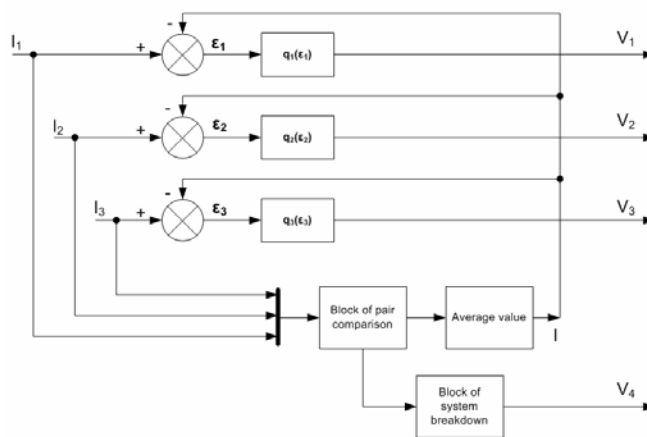
$\varepsilon_i = I_i - I$ difference between input and average signal
I_i – measured input value
V_i – output of kvorum element

Fig. 4. General scheme of quorum element

On the input of the quorum element are input variables (I_1, I_2, I_3), which can be measured identically (the same measuring method of the value on all inputs), or may represent signals obtained from the measurement of the variables by various methods.

Subsequently, they go through the block of pair comparison, where based on defined permitted error residue $\varepsilon_{i, pov}$, their comparison and evaluation. The output signal that exceeds the permissible error residues does not enter the next block, where the average value – I (1) is calculated.

$$I = \frac{\sum_{i}^{n} I_i}{n} \qquad (1)$$

where n – is a number of inputs, which do not exceed the permissible error residues.

This is in the form of negative feedback supplied to the input of block q_i. In the case, that in the pair comparison fails more inputs, the block signalizing total breakdown of the system is activated, this is indicated by the corresponding output V_4.

Block q_i has a characteristics (2) in the form:

$$q_i = \begin{cases} 1, for : \varepsilon_i \leq |\varepsilon_{i, pov}| \\ 0, for : \varepsilon_i > |\varepsilon_{i, pov}| \end{cases} \qquad (2)$$

serves as a member of the validation (Figure 5), where its output is a logical binary output *(0,1)*. This output (V_i – from V_1 to V_3) directly indicates if the input of quorum element represents measurement of error or not.
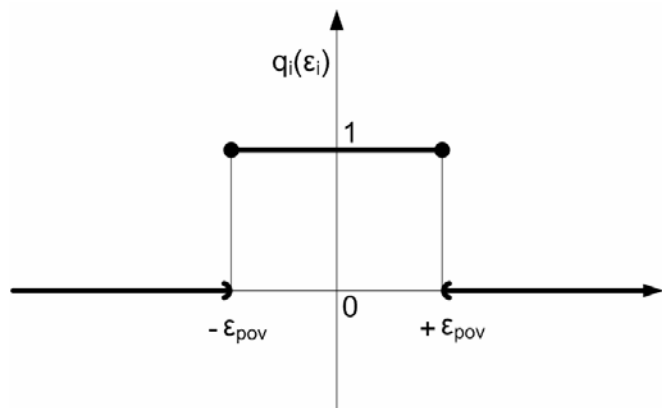
Fig. 5. Characteristics of signal $\varepsilon_{i,\,pov}$

REFERENCES

[1] J. HLAVIČKA, S. RACEK, T. BLAŽEK: Digital Fault tolerant systems. Praha : ČVUT, 1992. pp. 330 , ISBN 80-01-00852-5 (in Slovak language)
[2] J. HLAVIČKA: Diagnostics and reliability of digital systems - 2. edition. Praha : ČVUT, 1983. pp. 139 (in Slovak language)
[3] J. HLAVIČKA: Reliability and diagnostics. Praha : ČVUT, 1985. pp. 155, ISBN 80-01-00173-3 (in Slovak language)
[4] J. KABÁT: Usability a magnetic aura of small turbojet engine for the purposes of diagnostics. Thesis, LF TUKE, 2010, pp. 112 (in Slovak language)
[5] M. KREIDL: Diagnostics systems. Praha : ČVUT, 1997. pp. 314, ISBN 80-01-01670-6 (in Slovak language)
[6] T. LAZAR et. al.: Development tendencies and modeling of avionics systems : MoSR, Bratislava. 2000. pp. 160, ISBN 80-88842-26-3 (in Slovak language)
[7] T. LAZAR: provided unpublished materials available directly by author. (in Slovak language)
[8] L. MADARÁSZ, R. ANDOGA, L. FŐZŐ: Intelligent Technologies in Modeling and Control of Turbojet Engines. In: New Trends in Technologies: Control, Management, Computational Intelligence and Network Systems, Meng Joo Er (Ed.), Sciyo, Available on: http://www.intechopen.com/articles/show/title/intelligent-technologies-in-modeling-and-control-of-turbojet-engines, 2010. pp. 17 -38. ISBN: 978-953-307-213-5
[9] J. MODROVIČOVÁ: Magnetic aura study of small turbojet engine MPM 20 and possibilities of its usability in MPM 20 situational control. Thesis, FEI TUKE, 2010, pp. 124 (in Slovak language)

Based on the work [15], where inputs to the system (quorum element) represented three different methods of measurement of the operating speed MPM 20, allowed error residues (RPM) were defined, based on empirical experience. For q_i:

Optical sensor

$$q_1 = \begin{cases} 1, pre : \varepsilon_i \leq |2000| \\ 0, pre : \varepsilon_i > |2000| \end{cases} \tag{3}$$

Calculation of the operating speed based on changes in the magnetic field

$$q_2 = \begin{cases} 1, pre : \varepsilon_i \leq |500| \\ 0, pre : \varepsilon_i > |500| \end{cases} \tag{4}$$

Calculation based on a nonlinear model of turbojet engine

$$q_3 = \begin{cases} 1, pre : \varepsilon_i \leq |1500| \\ 0, pre : \varepsilon_i > |1500| \end{cases} \tag{5}$$

For these values the table of the block of pair comparison was constructed, see the following (Table 1).

TABLE I
THE TABLE OF THE BLOCK OF PAIR COMPARISON

Pair comparison of i and j inputs:	Allowed value of error $(\varepsilon_{i,pov} + \varepsilon_{j,pov})$[RPM]:
I_1 a I_2	2500
I_1 a I_3	3000
I_2 a I_3	1500

Towards Interactive Robots

Mária VIRČÍKOVÁ, Zlatko FEDOR

Center for Intelligent Technologies, Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice
www.ai—cit.sk

{maria.vircikova, zlatko.fedor}@tuke.sk

Abstract—**We present an interactive system in which a human user freely cooperates with humanoid robots. Our goal is the development of new robotic behaviors via the Interactive Evolutionary Computation (IEC). Firstly, robots learn motions observing their human teachers using Kinect—sensor device from Microsoft which provides full—body 3D motion capture. In the second phase, robots produce new motions and a human user evaluates them according to his preference. We also present a dialogue system which makes the communication with the robots natural. This process is leading to the emergence of creative and personalized robotic behaviors.**

Keywords—**Creative Design, Human—Robot Interaction, Interactive Evolution, Robotic Motion, Speech Recognition.**

> *"Creativity comes from looking for the unexpected and stepping outside our own experience."*
>
> Masaru Ibuka(1908—1997), co—founder of Sony

I. INTRODUCTORY NOTES

Richard Dawkins in his *Blind Watchmaker* [8] talks about a popular cliché which says that you cannot get out of computers any more than you put in. Other versions are that computers only do exactly what you tell them to, and that therefore computers are never creative. The cliché is true only in the crashingly trivial sense, the same sense in which Shakespeare never wrote anything except what his first schoolteacher taught him to write words.

In the 1956 Dartmouth Proposal for Artificial Intelligence [10], in a seven—point list of the *Artificial Intelligence problem*, creativity and randomness are specifically addressed: "Randomness and Creativity: A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of some randomness. The randomness must be guided by intuition to be efficient. In other words, the educated guess or the hunch includes controlled randomness in otherwise orderly thinking. "

The main purpose of designing an artificial creative systems is to support creative human thinking or, in higher level, to automate it. In research literature, many attempts trying to define creativity can be found. Boden [11] describes creativity as "a matter of using one's computational resources to explore, and sometimes to break out of, familiar conceptual spaces." Edmonds and Candy [12] define creativity as "a process toward achieving an outcome recognized as innovative." Kirton [13] differentiates between two creativity styles: adapters— like to deal with only a few ideas at a time
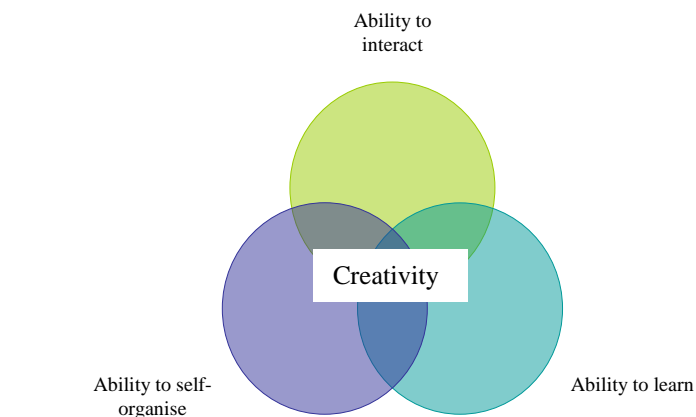

Fig. 1. Three requirements for creativity according to [19].

and innovators—who ignore or challenge systems, contribute radical proposals for change, are non—conforming and inefficient, and like ideas to proliferate. Cohen [14] argues that creativity applies to continuous change. More authors [14],[15],[16] agree that for human creativity it is important to create a stress—free, working environment with ample facilities for collaboration and exchange of ideas, and where making mistakes is allowed or even encouraged. AI researchers and cognitive scientists [17],[11] add that learning and interaction with the environment are important or necessary aspects of being creative. According to [18] learning may happen before the start of a design activity (provisional learning), in parallel with a design activity in progress and after completion of a design activity (retrospective learning). Authors [19] present a proposal for creativity requirements (as illustrates Figure I) that provide a basis for designing creative artificial systems: such system must be able to interact with its environment, learn and self—organise.

Much research has been done where IGAs are used to assist users in the development of creative ideas. GAs employs methods analogous to the processes of natural evolution in order to produce successive generations of software entities that are increasingly fit for their intended purpose. In GA, the crossover & the mutation are sources of variation.

Goldberg [21], under the consensus of *postmodern systems engineering,* formulated the "innovation intuition" within these two relations:

```
Selection + Mutation = Continual Improvement
   Selection + Recombination = Innovation
```

Having in mind the Goldberg´s concept of innovation intuition (more can be found in [21]), by the medium of the operators in genetic algorithms, a system makes up combinations (crossover or recombination), some chooses (selection), recognizes which are better (fitness function) and produces a change in the next generation of offspring (mutation).

Simply said, a recombination is a genetic operator which combines bits and pieces of parental solutions [21] and this way forms new, possibly better generation. One could say that there is nothing innovative, as the recombination, however, is not identical to any particular parent and will combine parent traits in a novel manner, but, always it is just a random shuffling of extant trails. Here, the mutation comes into play, modifying an individual. That way we can expect absolutely different solutions from that from previous generation.

Under creativity we mean producing something new or unexpected and valuable. A well known inhibitor of creative thinking is design fixation, where the suggestion of a particular solution makes it difficult to imagine other good solutions. Unlike many other computational search algorithms, EC methods work with populations of "fairly good" solutions. Therefore, there is less danger that creativity will be harmed by design fixation on one "best" solution.

The current leading topics in humanoid robotics are service and social robotics [3]—where the target is getting robots closer to people and their social needs. In our work we develop a robotic behavior where robot learns from human, observing him, in our case the robot learns new motions; and after he develops new ones through the process of interactive evolution. In our previous work [1] we constructed a system of generation of robotic motions based on the concept of Interactive Evolution. In this article we further study usage of Interactive Genetic Algorithm in production of robotic behaviors but we added learning—from— demonstration part preliminary to the evolution. Also the way in which a human communicates with the user was improved—the graphical user interface was changed for the dialogue system.

This paper is organized as follows: Section II describes our dialogue system designed for natural communication with the robot. Section III presents our approach where humanoid robots acquire some initial knowledge from human users— they learn some motion skills form humans via imitation learning. In the second part of this section we analyze in more detail an evolutionary approach for designing creative robotics behavior. Afterwards we present some of our experimental results and introduce our ideas for the future work.

II. DIALOGUE SYSTEM

New trends in research consider the robots ability to interact with people safely and in natural way [3]. We engage people in face—to—face interaction with a humanoid robot. We developed a dialogue system which allows the robot Nao recognize and learn new words through a simple dialogue with a human.

In [6] we implemented an incremental system for multi languages linguistic command recognition. It was realized in multi agent system MASS, based on client—server architecture. The system as well as documentation can be found on http://brain.fei.tuke.sk/. System was allowing a remote parallel learning of various commands, their consecutive identification and robot dog AIBO control. Now we adapted and tested this system with the humanoid robots.

We examine Mel Frequency Cepstral Coefficients (MFCCs) - the dominant features used for speech recognition. We use this method for preprocessing of speech signals, because it copies the process of pre-processing speech in the human brain. After the reduction of the information flow and creation of decorrelated Mel-spectral vectors, we use a method for classify words called Dynamic Time Warping (DTW). DTW is an algorithm for measuring similarity between two sequences which may vary in time or speed. In the experiments we evaluated the method MFCC importance compared to unpreprocessed signal for the classification of isolated words spoken by several speakers.

Thanks to our dialogue system inexperienced human users evaluate the behaviors of robots simply, talking to them— giving them commands and which can robots understand. The key contribution of the "multi agent system" approach is that when one robot learns a user´s verbal command, also "all robots in the family" will know it. For better understanding, a demonstration of our approach can be seen on website http://www.vimeo.com/22578478.

III. LEARNING MOTION PATTERNS VIA INTERACTIVE EVOLUTION

In our previous research [1] we developed a multi—robot system capable of evolution of motions using an Interactive Evolutionary Computation [5]. This technique optimizes systems based on subjective human evaluation. Simply said, it is an Evolutionary Algorithm in which the fitness function is replaced by human. Users evaluate different sequences of robotic movements and give them marks according to their preferences. This evaluation is a graphical user interface where a user selected mark from 1 to 5 depending on how he liked the presented motion. Implementing the dialogue system described in Section II, this part turns into a free interaction with real robots where a user evaluates robotic behaviors verbally. After user evaluates the first generation of motions of the robots, the Evolutionary Algorithm produces the next generation.

If the evolution starts from random movements, the process of interaction until satisfactory results is long so we implemented learning—from—human stage in the initial phase of designing the motion. An agent learns a behavior from observing the execution of that behavior by a teacher (imitation learning)—an inexperienced human user interactively cooperates with the system and this is the way how the robot acquires knowledge.

Authors [7] formulated a statement: Imitation takes place when an agent learns a behavior from observing the execution of that behavior by a teacher. They also enumerated the advantages a learning mechanism should give: adaptation, efficient communication, compatibility with other learning mechanisms, efficient learning and thus, learning by imitation seems to be one of the major components of general intelligent behavior. Imitation encompasses a set of different competences such as recognizing other people's actions, recognizing the goal of a particular action and the objects and/or subjects involved.
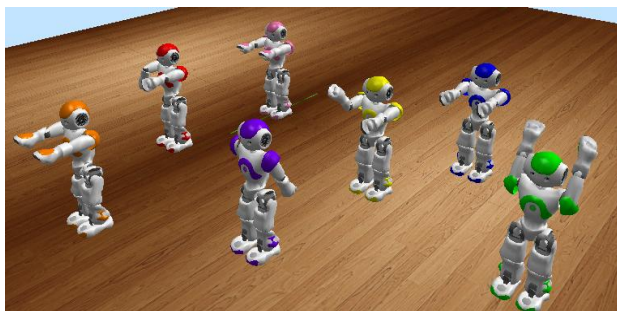
Fig. 2. Simulation of our System of Evolution of Robotic Motions, in which user evaluates each of them and posterior motions are evolving in accordance to his evaluations.

Approach of learning tasks from demonstrations according to [8] is a convenient mechanism to customize and train a robot by transferring task related knowledge from a user to a robot. This avoids the time—consuming and complex process of manual programming. The way in which the user interacts with a robot during a demonstration plays a vital role in terms of how effectively and accurately the user is able to provide a demonstration.

The key question was how to map a teacher action to an agent action. These agent actions in our experiments are the motor capabilities of the robots. We mimic user movements using Microsoft Kinect sensor device, which provides between others full—body 3D motion capture capability. The hardware emits and infrared light which rebounds off object in the room. After processing the rebounding light, the system differentiates between people and other objects. The system gives the body a virtual skeleton with moving joints so the machine can follow user movements. From programmer´s point of view, rather than having access to the pixels of a flat image, or even the depth information from the raw Kinect data, we have a description of users as a series of joints in space. For many applications, especially gesture-based interfaces and motion capture, this is exactly the data which we need to get started.
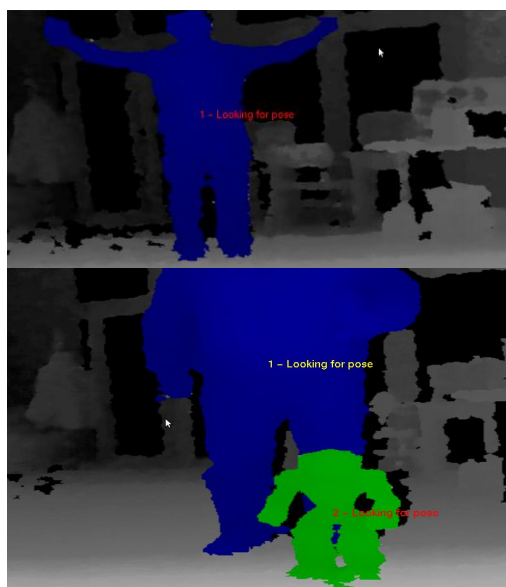
Fig. 3. How the sensor device Kinect maps a person (blue) and the Nao robot (green).

TABLE I
MOTION RANGE

Motion	Range(degrees)
Head Yaw – twist	-120 / 120
Head Pitch – front & back (Y)	-39 / 30
Left Shoulder Pitch – front & back (Y)	-120 / 120
Left Shoulder Roll Left – right & left (Z)	0/95
Left Elbow Roll – twist (X)	-90 / 0
Left Elbow Yaw (Z)	-120/120
Left Wrist Yaw – twist (X)	-105/105
Left Hand	Open/close
Left Hip Yaw Pitch – twist (Z45°)	-44/68
Left Hip Pitch – front and back (Y)	-104.5/28.5
Left Hip Roll – right & left (X)	-45/25
Left Knee Pitch	-5/125
Left Ankle Pitch – front & back (Y)	-70.5/54
Left Ankle Roll – right & left (X)	-45/25
Right Hip Yaw Pitch – twist (Z45°)	-68/44
Right Hip Pitch – front and back (Y)	-105.5/28.5
Right Hip Roll – right & left (X)	-45/25
Right Knee Pitch (Y)	-5/125
Right Ankle Pitch – front & back (Y)	-70.5/54
Right Ankle Roll – right & left (X)	-25/45
Right Shoulder Pitch – front & back (Y)	-120/120
Right Shoulder Roll Right – right & left (Z)	-95/0
Right Elbow Roll – twist (X)	0/90
Right Elbow Yaw (Z)	-120/120
Right Wrist Yaw – twist (X)	-105/105
Right Hand	Open/close

Joint names of robot and possible motions – range in degrees.

We work with humanoid robot Nao developed by Aldebaran Robotics [2]. The robot has a human—like appearance and various sensors for interacting with humans. In community of artificial intelligence, by humanoid we mean not only a robot that has the physical appearance of a human, but also has a fairly advanced artificial intelligence, allowing it to reasonably approximate human behavior and interactions. As our robot increasingly resemble human in shape, we map his servos (presented in Table I) to the points of human body. This way we model different robot´s motor skills via its interaction with several human users—we personalize robot´s behaviors according to the interaction of the specific human user.

Interactive Evolutionary Computation, introduced by Takagi [5] is an optimization method that adopts EC among system optimization based on subjective human evaluation. IGA can reflect personal preference because it percepts fitness directly from user instead of computing using some function. As said before, it is simply an EC technique whose fitness function is replaced by a human user or better said; it uses human's response as a fitness value. The user acts as critic of the robotic behaviors. We can say that the IEC is a technology that embeds human preference, intuition, emotion, psychological aspects, or a more general term, KANSEI, in the target system.

To create an initial population, as described above, we first teach the robot some initial movements using sensor Kinect. The algorithm calculates positions of each joint of human and maps it to the robot´s joint. Each phenotype represents one joint of Nao and his position in angles. The value of fitness in interactive evolutionary algorithm is obtained from user of the system. Users evaluate robotic motions and "give them marks" according to their preferences.
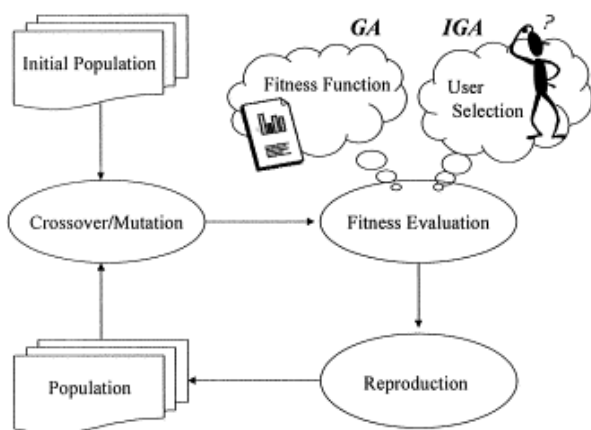
Fig. 4. Process of Interactive Evolutionary Computation by Takagi [5], which uses human's response as a fitness value.

We use the tournament selection method. The winner of each tournament is the one with the highest fitness value and is selected for crossover.

The crossover is an important source of variation in genetic algorithm. We used a uniform crossover where the bits that are exchanged depend on the masks.

The mutation is another source of variation in genetic algorithms. The probability of mutation of every individual is 80 percent but in a very close neighborhood of each bit (less than 10 percent), so the changes are „small".

IV. EXPERIMENTAL RESULTS

We have conducted several evaluation experiments so far. In the first experiment, a group of twenty students with different backgrounds and preferences used our system to make robotic movements whatever they want. Most of the subjects found our system performance satisfactory. A user can build a robot behavior without any programming knowledge and then the evolutionary algorithm helps user to discover new possibilities of motions. In another experiment, we gave users the theme to teach robot some behavior pattern, such as "ballet dance", and then make them evolve dance by our system.

Creating an initial population using sensor Kinect helped to the reduction of GA population and we got a positive feedback from users. A couple of generated dances in these experiments are available to see on our website[23].

To evaluate the effectiveness of this method, however, we need to analyze experimental results more quantitatively, which will be our future work.

V. CONCLUSION & FUTURE WORK

We proposed a system which at first learns a set of motion behaviors from human and then creates new behaviors exploring the search space of the evolutionary algorithm During this process the human user evaluates his achievements and so it produces "better" behaviors from the point of view of the user.

The paper briefly analyzed the connection between evolutionary algorithms and creativity, or in the lower level, innovation. We think understanding innovation in mathematical—computational terms could be a big contribution in the design of systems and also could help in the better understanding of human creativity, which we consider a significant part of intelligence.

In the future work, our goal is for Nao to acquire its own mental model of people. We want to extract the information about the own preferences of human during his evaluation of the behavior of the robot during the IEC and make this process autonomous. We also plan to extend the set of motions learnt by the robot in his interaction with a user.

We conclude this paper with words of the British author Bulwer—Lytton [22] who said:

"Invention is nothing more than a fine derivation from, or enlargement on a fine model.. .. Imitation, if noble and general, insures the best hope of originality."

REFERENCES

[1] Vircikova, M.: Artificial Intelligence in Humanoid Systems, Diploma Thesis, 2010
[2] Gouaillier, D., Hugel, V. et. al.: The Nao Humanoid: A Combination Of Performance And Affordability. In: IEEE Transactions On Robotics, 2008
[3] Garcia, E., Jimenez, M., A. et. al.: The Evolution Of Robotics Research, From Industrial Robotics To Field And Service Robotics. In: IEEE Robotics And Automation Magazine, 2007
[4] Fong, T., Nourbakhsh, I., Dautenhahn., K.: A Survey of Socially Interactive Robots: Concepts, Design and Applications, Technical report no. Cmu—Ri—Tr—02—29, Carnegie Mellon University, 2002
[5] Takagi, H.: Interactive Evolutionary Computation. Cooperation Of Computational Intelligence and Human Kansei. In: Proceedings of the IEEE, Vol. 89, Issue 9, pp. 1275 — 1296, ISSN: 0018—9219, 2001
[6] Fedor, Z.—Sincak, P.: AIBO talking procedure in multi languages based on incremental learning approach , In: SAMI 2009 : 7th International Symposium on Applied Machine Intelligence and Informatics : January 30-31, 2009, Herľany, Slovakia. - S.l. : IEEE, 2009. - ISBN 978-1-4244-3802-0. – pp. 141-14,5 2009.
[7] Berthouze, L.—Bakker, P.—Kuniyoshi, Y.: Learning of Oculo—Motor Control: A prelude to Robotic Imitation. In: Proceedings of the 1996 Lee/RSJ International Conference on Intelligent Robots and Systems, 1996
[8] Koenig, N.—Takayama, L.—Mataric, M:: 2010 Special Issue: Communication and Knowledge Sharing in Human—Robot Interaction and Learning from Demonstration. In: Journal Neural Networks, Vol. 23, Issue 9, Elsevier Science ltd. Oxford, 2010
[9] Dawkins, R.: The Blind Watchmaker, 1986
[10] 1956 Dartmouth Proposal for Artificial Intelligence, 1956
[11] Boden, M.A.: Creativity and artificial intelligence. Artificial Intelligence, 103(1—2), 347–356, 1998
[12] Edmonds, E., & Candy, L.: Creativity, art practice, and knowledge. Communications of the ACM, 45(10), 91–9, 2002
[13] Kirton, M.J.: Adaptors and innovators: a description and measure. Journal of Applied Psychology, 61(5), 622–629. Reprinted in: Creativity Assessment (Puccio, G.J., & Murdock, M.C., Eds.), 1999, pp. 121–132. Hadley: The Creative Education Foundation Press
[14] Cohen, H.: A self—defining game for one player: on the nature of creativity and the possibility of creative computer programs. Leonardo, 35(1), 59–64, 2002
[15] Weatherall, A.: Creative problem solving using GroupSystems. Proc. Thirty—first Annual Hawaii Int. Conf. on System Sciences, Vol. 1, pp. 588–595, 1998
[16] Greene, S.L.: Characteristics of applications that support creativity. Communications of the ACM, 45(10), 100–104, 2002
[17] Simon, H. A.: The Sciences of the Artificial. Cambridge: MIT Pres, 1969
[18] Sim, S.K., & Duffy, A.H.B.: Evaluating a model of learning in design using protocol analysis. Proc. Sixth Int. Conf. on Artificial Intelligence in Design (AID '00) (Gero, J.S., Ed.), pp. 455–477. Dordrecht: KluIr Academic Publishers, 2000
[19] Van Langen, P.—Wijngaards, N.—Brazie, F: Towards Designing Creative Artificial Systems
[20] Norman, K. L.: Cyberpsychology. An introduction to Human—Computer Interaction, Cambridge University Press, New York, 2008
[21] Goldberg, D. E: The design of innovation: Lessons from and for competent genetic algorithms, Boston, MA: Kweuwer Academic, 2002
[22] Asimov, I., & Shulman, J. A. (Eds.) Isaac Asimov's book of science and nature quotations. New York, NY: Weidenfeld & Nicolson, 1988
[23] url: http://neuron.tuke.sk/~maria.vircik/danceproject/

Verification of Control Algorithms with DDE Communication on Real Hydraulic System

Štefan JAJČIŠIN

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

stefan.jajcisin@tuke.sk

Abstract—**The paper deals with a real hydraulic system modeling and control by designed digital PID control algorithms in Matlab with using DDE communication. A mathematic-physical description of real hydraulic system is established in a form of nonlinear differential equations on the basis of analytic identification with known physical principles. Also simulation model, created in Simulink, based on nonlinear differential equations for designed PID control algorithms verification purpose is introduced in the paper. The design of digital PID control algorithms is briefly mentioned in the next section of the paper. A flow chart diagram of implementation of main control program with DDE communication between the real system and a computer, on which control algorithms are running, is a part of the paper, too. In the end of the paper results of real hydraulic system control by designed digital PID control algorithms with DDE communication are presented.**

Keywords—**communication protocol DDE, PLC, digital PID control, real hydraulic system, simulation language Matlab.**

I. INTRODUCTION

The paper deals with modeling of real hydraulic system, which serves as an educational model for identification and control algorithms verification at the Department of Cybernetics and Artificial Intelligence. The paper is also engaged in control of real hydraulic system with designed digital PID control algorithms in Matlab with using DDE communication.

The first part of the paper is devoted to real hydraulic system mathematic-physical description, which was derived in form of nonlinear differential equations on the basis of analytic identification with using known physical principles. Moreover, a hardware structure of real hydraulic system and a way of its connection to the computer is presented. Also simulation model created in Simulink by implementation of nonlinear differential equations for purpose of closed loop control structure with designed PID control algorithms analysis is introduced in the paper, too.

In the next section of the paper a design of digital PID control algorithms is presented. Also a program implementation of DDE communication protocol in the framework of control algorithm running in Matlab is mentioned.

The paper contains results of real hydraulic system control by designed digital PID control algorithms with DDE communication verification in form of control action and controlled output time responses.

II. REAL HYDRAULIC SYSTEM

The real hydraulic system is one of laboratory systems, which serve as educational models for identification and control algorithms verification at the Department of Cybernetics and Artificial Intelligence. It is composed of two cylindrical tanks, which are connected by a tube. Tanks are placed one above another and they constitute a system of tanks without mutual interactions. A liquid is fetched into the first tank from a bathtub through an inlet tube by a pump [1]. It flows out from the second tank back to the bathtub under the thumb of hydrostatic pressure. It is possible to change an intersection size of outlets in both tanks by throttling screws. There are four different intersection sizes.

A schematic illustration of real hydraulic system is depicted in Fig. 1, whereby particular physical parameters are:

S - intersection of tanks,
S_{v1}, S_{v2} - intersection of outlets of both tanks,
h_{max} - height of tanks (maximal liquid level),
$f_m(t)$ - pump's motor frequency,
$h_1(t)$, $h_2(t)$ - current levels of liquid in both tanks.

Sensors, which scan the current liquid level in both tanks are marked as S_{n1} and S_{n2}.
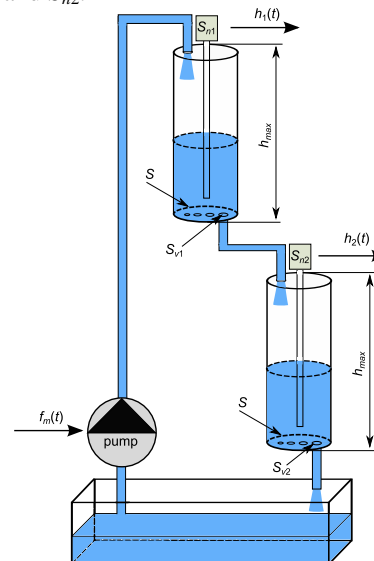

Fig. 1 Hydraulic system – physical principle

The frequency $f_m(t)$ constitutes the only input of laboratory model in term of systemic approach. The frequency $f_m(t)$ is generated by frequency drive [2] according to voltage $U(t)$, which is connected to frequency drive's input from analog

card of PLC [3] in the range 0 – 10V. Thus, it is possible to consider the direct value of voltage $U(t)$ as laboratory model's input. There is a linear dependence between the voltage $U(t)$ and frequency $f_m(t)$.

Outputs of laboratory model are represented by current levels of liquid in both tanks $h_1(t)$ and $h_2(t)$, which are measured by sensors [4]. However, the aim of control is to ensure required value of liquid level in the second tank h_{2ref}. Therefore we restrict ourselves to a SISO system control, where the value $h_1(t)$ serves only for demands of PLC program, which handles limit values of particular quantities and emergency conditions.

The change of intersection size of the second tank outlets ΔS_{v2} constitutes a disturbance.

The systemic view of real hydraulic system is depicted in Fig. 2.
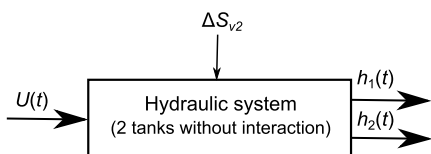
Fig. 2 Hydraulic system – systemic view

A. Structure of real hydraulic system

A complete structure of real hydraulic system is depicted in Fig 3. The real hydraulic system is directly joined to input/output analog card of PLC. The program in PLC, which was created in RSLogix environment [5], provides reading and writing values of particular quantities to the PLC card. Moreover, it safeguards system protection from emergency conditions.

Concurrently, PLC is connected to a personal computer by serial or Ethernet interface. The communication between PLC and PC is carried out in such a way. It is necessary to install software providing DDE server functions to PC. In this case it is RSLinx [6], because PLC is product of Rockwell Automation company [3].
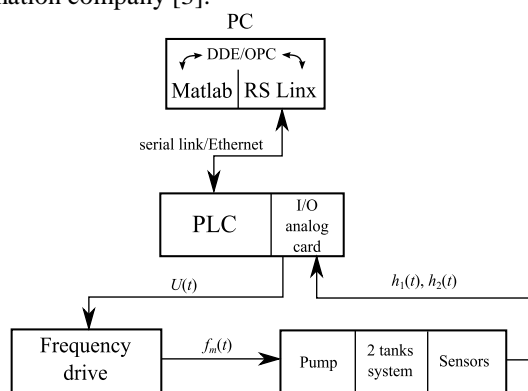
Fig. 3 Structure of real hydraulic system with connection to PC

Mentioned structure of real hydraulic system makes possible to carry out control in two different ways. The simplest way is to implement designed digital PID control algorithms directly to the PLC, which allows to leave out PC of structure in Fig. 3. Creating communication between PLC and PC becomes unimportant in this case. The second approach is based on the control algorithms running in Matlab in PC. This approach is more difficult in term of creating communication. On the other side, it allows to use many advantages of Matlab. I have decided for the second way because I would like to verify predictive control algorithms on

the real hydraulic system, which are based on quadratic programming with using *quadprog* function from Optimization toolbox of Matlab in future.

B. Mathematic-physical description of hydraulic system

It is possible to derive nonlinear differential equations by analytic identification with using known physical principles, such as continuity relation or Torricelli's formula:

$$\begin{aligned} \frac{dh_1(t)}{dt} &= \frac{1}{S}\left(k_p U(t) - S_{v1}\sqrt{2gh_1(t)}\right) \\ \frac{dh_2(t)}{dt} &= \frac{1}{S}\left(S_{v1}\sqrt{2gh_1(t)} - S_{v2}\sqrt{2gh_2(t)}\right) \end{aligned}, \tag{1}$$

which describe hydraulic system dynamics, whereby g is acceleration of gravity. Pump's value k_p was computed from experimental measurements.

It is important to note that it is necessary to establish a linear model, which would represent the dynamics of hydraulic system at properly chosen operating point for digital PID control algorithms design.

C. Nonlinear simulation model of hydraulic system

A nonlinear simulation model of hydraulic system, which is depicted in Fig. 4, was created in Simulink on the basis of nonlinear differential equations (1). The simulation model was used in the analysis of control closed loop with designed digital PID control algorithm dynamics.
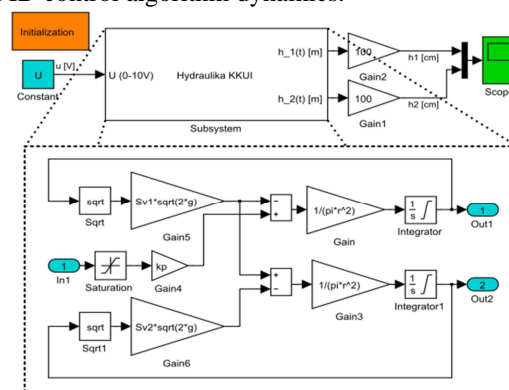

Fig. 4 Nonlinear simulation model of hydraulic system

III. DIGITAL PID CONTROL ALGORITHMS DESIGN

Digital PID control algorithms design starts from the linear model of hydraulic system, which was obtained by nonlinear differential equations (1) linearization in operating point:
$U_0 = 4V$, $h_{10} = 0.16m$, $h_{20} = 0.158m$, whereby $h_{max} = 0.3m$.
The linear model has a form of transfer function

$$F_s = \frac{H_2(s)}{U(s)} = \frac{K}{\left(a_2 s^2 + a_1 s + 1\right)}, \tag{2}$$

where K is a static gain and a_i are denominator's elements.
It is possible to compute coefficients of continuous PID

$$F_R = P + \frac{I}{s} + Ds, \tag{3}$$

where P, I and D is proportional, derivative and integration gain.
The optimal module method and the standard form method by Butterworth were used for P, I and D coefficients design purpose. The latter method have been already used in digital PID control algorithms verification on the real system Ball and Plate in [7].
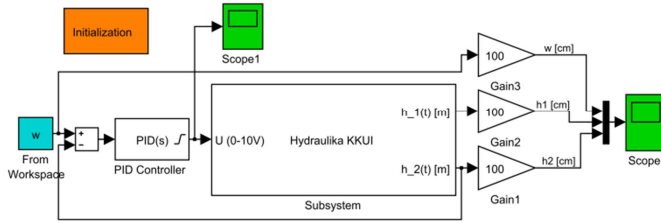
Fig. 5 Closed loop structure for nonlinear simulation model of hydraulic system control

It was found by the analysis of dynamics of the closed loop structure with designed PID control algorithm and simulation model (Fig. 5) that the derivative part in the control algorithm causes very fast increase of liquid level in the first tank. Thus, the maximal liquid level overflow would come into being and PLC program would automatically stop the frequency drive in real system control. For that reason, the digital PI control algorithm, which derivative part is zero, was designed and applied in the real system control structure.

In the case of the optimal module method a matrix form was used for computing P and I coefficients:

$$\begin{bmatrix} a_1 & -1 \\ a_3 & -a_2 \end{bmatrix} \begin{bmatrix} I \\ P \end{bmatrix} = \frac{1}{2K} \begin{bmatrix} 1 \\ -a_1^2 + 2a_2 \end{bmatrix}, \qquad (4)$$

where coefficients a_i are elements of transfer function's denominator. In the second case the standard form of 3rd order characteristic polynomial of closed loop according to Butterworth was used.

Coefficients of digital PID control algorithm q_i were computed from known values of continuous PID coefficients on the basis of equations mentioned in [8]. A formula for computing the control action, which is implemented in control closed loop for the real hydraulic system control, has form

$$u(k) = u(k-1) + q_0 e(k) + q_1 e(k-1) + q_2 e(k-2), \qquad (5)$$

where $e(k)$ constitutes a deviation between reference trajectory $w(k)$ and current system output value $y(k) = h_2(t)$.

IV. PID CONTROL ALGORITHMS VERIFICATION ON REAL HYDRAULIC SYSTEM

In this case, where aim of the control was fulfilled in simulation model control, designed PID control algorithm was implemented to the main program for real hydraulic system control in Matlab. Matlab functions of DDE protocol and RSLinx tools were used for communication with PLC. It is possible to read data from real system and subsequently change the control action value, which is computed by control algorithm in PC in such a way.

The basic principle of main program with DDE communication and digital PID control algorithm is visible in a flow chart diagram, which is depicted in Fig. 6. There are also Matlab commands providing DDE communication with PLC mentioned in the flow chart diagram.

Results of the real hydraulic system control using digital PI control algorithm, which coefficients were designed by the optimal module method are depicted in Fig. 7. The reference trajectory is a constant function with desired value, whereby another second tank's outlet was opened for a short while in time $t = 580$s. So, a disturbance was incorporated to control process in such a way.
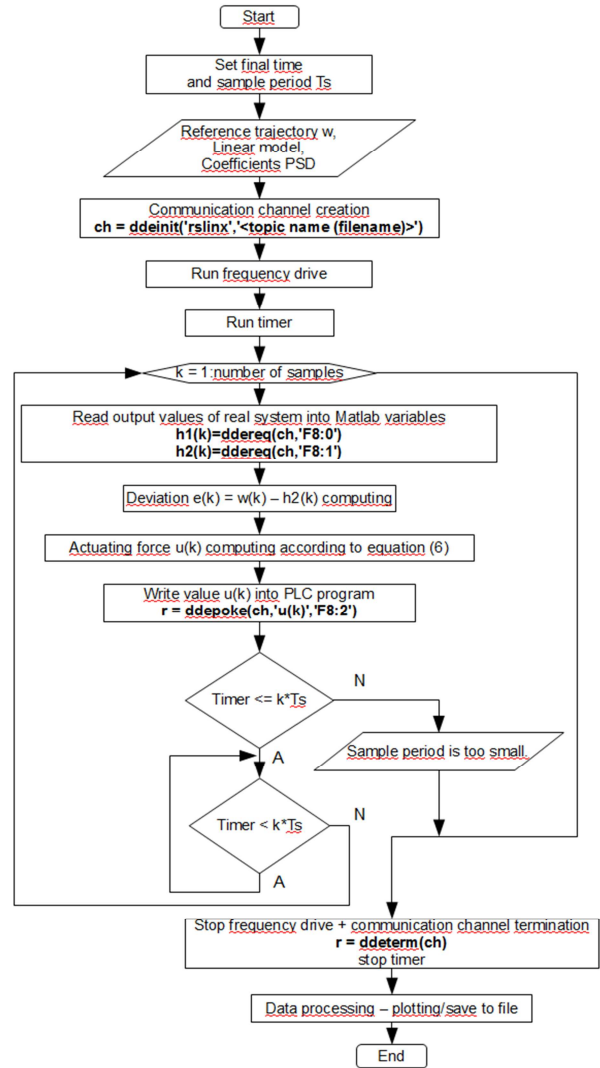
Fig. 6 Flow chart diagram of program with control algorithm using DDE communication
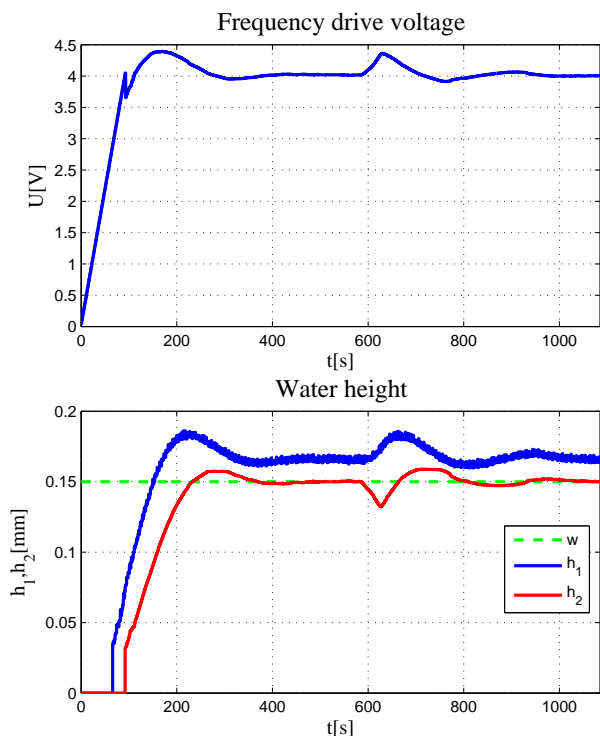

Fig. 7 Results of the real hydraulic system control with digital PI control algorithm designed by the optimal module method

In Fig. 8 results of the real hydraulic system control with digital PI control algorithm designed by the standard form method according to Butterworth are visible. In this case the reference trajectory was variable and control process was not affected by any disturbances.
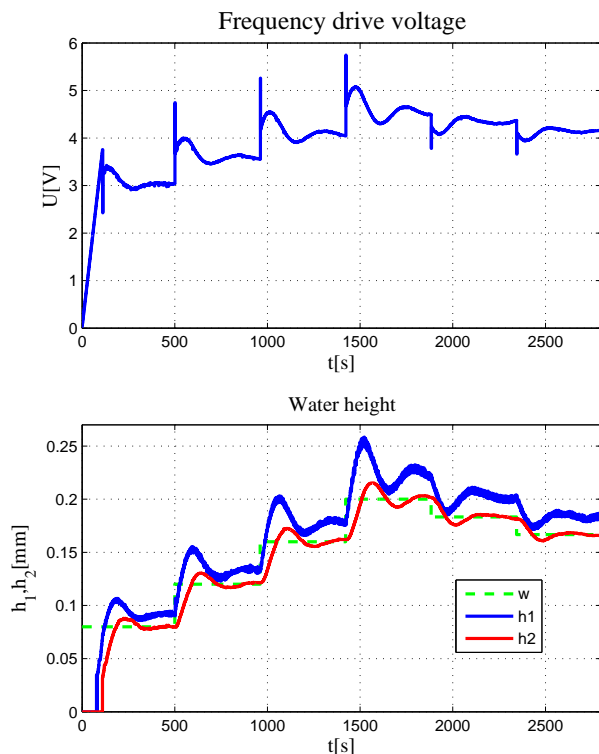

Fig. 8 Results of the real hydraulic system control with digital PI control algorithm designed by the standard form method according to Butterworth

By Fig. 7 and Fig. 8 comparison it is clear that using digital PI control algorithm designed on the basis of the optimal module method causes a smaller overshoot of controlled quantity with the step change of reference trajectory. This fact was confirmed by comparison of real hydraulic system control results, where digital PI control algorithms designed by both methods were used and the reference trajectory was the constant function.

Essentially it can be pointed out that real hydraulic system control with digital PID control algorithms enables to fulfill aims of control. However, it is necessary to note that the maximal liquid level overflow in the first tank does not occur only when the derivative part is close to zero. The digital PI control algorithm, designed on the basis of the linear model for properly chosen operating point, is robust enough to change the desired value of liquid level in the second tank. However, in selection the range of reference trajectory it is necessary to take fact that the liquid level in the first tank is higher than in the second tank into consideration.

V. CONCLUSION

The real hydraulic system, which constitutes an educational model for identification and control algorithms verification at the Department of Cybernetics and Artificial Intelligence was introduced in the paper. Its mathematical description in form of nonlinear differential equations, which served as the base for creating the nonlinear simulation model in Simulink was mentioned.

The main goal of this paper was to present the structure of the real hydraulic system connection to the PC through the PLC and possibility to carry out real system control by control algorithms, which are programmed and running in Matlab. Commands providing DDE communication between PLC and Matlab were introduced in the main program flow chart diagram.

Results of the real hydraulic system control with digital PI control algorithms were presented in the end of paper. Besides using PID control algorithms my aim is also to verify algorithms of modern control theory on this system, for instance linear quadratic algorithm based on the state space model or predictive control algorithms, where the linear model of system, created by nonlinear differential equations linearization, will be used as a predictor. However, I also plan to obtain the linear model of system as a linear regression ARX model by experimental identification of real system.

My next aim is to use real hydraulic system in nonlinear predictive control algorithms with a neural network. The neural network will be trained on the basis of measured data from real hydraulic system and it will serve as nonlinear predictor.

REFERENCES

[1] Bia-Verder, *SPRING diaphragm concept* (in Czech), technical documentation. Available on the internet: http://bia-verder.verder.ams3.footsteps.nl/Cerpadla/Davkovaci_cerpadla/Mechanicka/Seko_SPRING_MS.

[2] Rockwell Automation, *Allen-Bradley PowerFlex 40 Adjustable Frequency AC Drive*, User Manual. Available on the internet: http://www.ab.com/support/abdrives/webupdate/literature.html.

[3] Rockwell Automation, *MicroLogix 1200 Programmable Controllers*, User Manual. Available on the internet: http://literature.rockwellautomation.com/idc/groups/literature/documents/um/1762-um001_-en-p.pdf.

[4] Dinel, *Capacitance level meters CLM-36* (in Czech), User manual. Available on the internet: http://www.dinel.cz/uploads/pdf/090916081606-clm-36-nav-cz.pdf.

[5] Rockwell Automation, *RSLogix 500,* Product Profile. Available on the internet: http://literature.rockwellautomation.com/idc/groups/literature/documents/pp/lg500-pp001_-en-e.pdf.

[6] Rockwell Automation, *RSLinx Classic,* Getting Results Guide. Available on the internet: http://literature.rockwellautomation.com/idc/groups/literature/documents/gr/linx-gr001_-en-e.pdf.

[7] A. Jadlovská, Š. Jajčišin, R. Lonščák, "Modelling and PID Control Design of Nonlinear Educational Model Ball & Plate, *Proceedings of the 17th International Conference on Process Control '09*, Štrbské Pleso, High Tatras, Slovakia, 2009, pp. 475–483, ISBN 978-80-227-3081-5.

[8] L. Harsányi, J. Murgaš, D. Rosinová, A. Kozáková, "Automatic Control Theory (in Slovak)", STU Bratislava, 1998.

Visualized database of laboratory instrumentations

[1]Daniel BENETÍN, [2]Milan GUZAN

[1]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic
[2]Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

[1]dbenetin@gmail.com, [2]milan.guzan@tuke.sk

Abstract— **The rise of Web applications, XHTML documents, CSS, Java Script and open source software is becoming a phenomenon of modern information technology, which is used by many successful and quality software products. This article describes the practical deployment of open source products in the business sector and demonstrates that open source software offers advantages in many areas such as cost savings, and improve security and stability of the IT environment.**

Keywords—**CSS, Database, HTML, Java, JQuery, MySQL, PHP.**

I. INTRODUCTION

This article describes the techniques of processing and review visualized database instrumentation laboratory. This article provides information on how, when and why to use the tools.

As a basis for a database uses the MySQL open source project with a clear location data, programming environment provides access and modify the PHP programming language and visualization and graphical display of caring XHTML, Java script library, JQuery and CSS Cascading Style Sheets.

Database after certain adaptations can also be used for cataloging electronic components or measuring instruments purchased with support of the "Centre of Excellence of the Integrated Research & Exploitation the Advanced Materials and Technologies in the Automotive Electronics".

II. SYSTEM REQUIREMENTS

A. Harware part of system

Web database on your computer needs to run - server. This computer must be connected either to the local network, or to the Internet, depending on whether we require access to data only from local network or from anywhere with an Internet connection. Connecting to a network server for centralized access to visualization laboratories, with the possibility of data is always current. As a test server was used a classic desktop PC configuration with Intel Quad Core Q9400, 4 GB RAM.

B. Software part of system

All the techniques used for visualized database are usable in current operating systems such as Unix, Linux and Microsoft Windows. As operating system was selected Ubuntu linux distribution 10.10 64 bit, for its simplicity, easy availability and easy installation program using the packaging system and security associated with frequent updates of system and packages.

Software requirements:
- Apache2
- PHP5
- PHP5-devel
- PHP5-pear
- PHP5-MYSQL
- Pdo
- Pdo_mysql
- MySQL5 server
- Java (client station)
- Web browser (client station)

III. DATABASE STRUCTURE

Mysql database was designed around the needs, values that we want to store the tables on the basis of descriptions and device parameters as those in it will be found.

Abbreviated example of the parameters described of the equipment located in laboratories.

TABLE I
SAMPLE PARAMETERS OF MEASURING INSTRUMENTS - AN EXTRACT FROM THE DATABASE

Name:	S/N:	Range:	Dept. mark:	Box:
W	9376007	240,360 V/10A	57/69	W1
W	239375	120,240V/10A	139/69	W1
W	239396	120,240V/10A	140/69	W1
W	1494587	600V/1,2.5,5A	1510/67	W1
Ferometer	1740005		30/71	W2

Data model of the mysql tables looks like this.

Fig. 1 Physical data model

There are not links between tables because the tables manage different functionality of the system and it is user login and device database so do not need additional tables.

Example, a system of images where one device is always one picture to one unit works on the basis of the name of pictures by id of device. Additional tables would be the case if it had device include more pictures.

IV. PROJECT PROPERATION

Preparation of the project to create application was consisted of digitization of the necessary data, captured all the devices and their parameters and describing position in the cabinets.

All values of the instruments and laboratory equipment I ordered in a summary tables (MS Excel, OpenOffice Calc - data from such a table can be easily prepared with assistance of developed php script to import directly into MySQL database), including their location and then create a photo documentation of the laboratory, cabinets and equipment in cabinets is shown in Fig. 1.

Fig. 2 View of the cabinets in the laboratory, where equipment is stored.

Then was created several images of visualized cabinets – animation of opening and closing the doors. Using Javascript to create the necessary 'hover' effect - opening or closing the door when you move the mouse over the box. Icons in the cabinet are used as miniature devices.

After passing the thumbnail opens a bubble with a picture and description of selected device. Door opening animation is shown in Fig. 2.
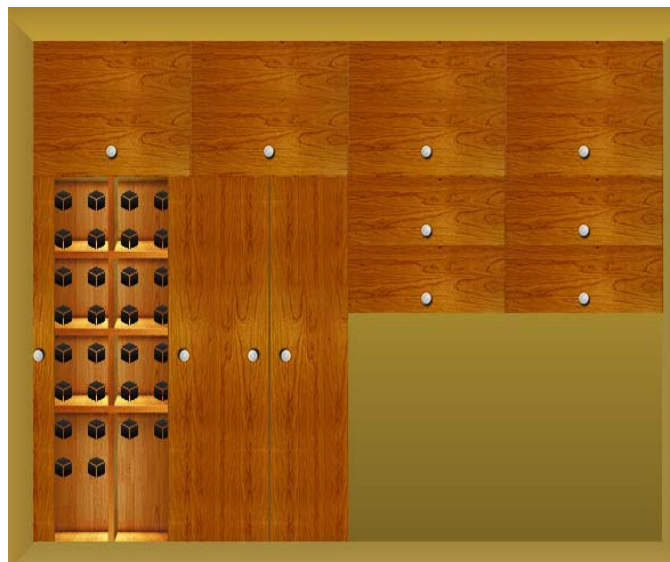
Fig. 3 Hover effect – opening and closing

V. SYSTEM CAPABILITIES

For the protection and safety system is required to log into the application with login and password as the data in the database may be sensitive (laboratory equipment are expensive and do not need to unauthorized person knew where the closet is located).

Administrator can add users to the database system. User editation menu is shown in Fig. 3

Fig. 4 User administration table

The system is divided into two user levels, first level of user in the database can only search and view specific device where it is stored.

The second level is administrator and his rights is moreover, he can make the database more work, for example to add and edit equipments, import and configure other laboratories, configure the layout of boxes in the laboratory, upload pictures and descriptions of facilities and also be able to move devices in cabinets as shown in Fig. 4.

Detailed infomation about the device are shown in tabular form for each device. This information includes all elements of the items from the database arranged in tabular form.

Fig. 5 Editation of apparatus

He can also add and delete users of the database system. Displaying items in the databae system contains the functionality of filtering these items using specification search parameters and thereby users will be able to get the information required by the device. Filtering apllied cumulatively with the possibility of applying 0 to n filters in conjuction so that the number of items returned declined until the user is required to finalize the device. The components of the database to whitch filtering apllies are specified according to the importance of the elements. Filter has a function to export selected data to the format csv.

Fig. 6 Visualized apparatus

VI. PROGRAMMIMG

The basic building block of the entire application is PHP Zend framework linked to mysql database.

Zend Framework is a PHP library for developing applications in PHP. Contains many components that allow PHP applications easier to develop a sustainable code for future modifications and improvements. It has extensive documentation, large community, applications written using the Zend Framework to develop simple and allows rapid development and expansion.

The core system consists of individual modules, with which we use zend_controller – control entries in the database, zend_view for viewing data, zend_db for configuration login parameters into mysql, zend_config, zend_filter for filtering viewed content, zend_validate, zend_acl for manipulation with user rights, zend_auth for user authentification and zend_registry.

Components forming the core of Zend Framework using the so-called MCV (Model-controller-view) system for production applications where appearance is a separate web site from logic and controller. The main part of MCV therefore constitute: Zend_Controller (Controller), Zend_View (View) a Zend_Db (Model).

Software development was aimed to best user friendly display and orientation in boxes, quick loading content of cabinets, which was provided by adjusting images for greater compression for faster preview thumbnails. To simplify the php code and the possibility of further expansion by other programmers, the source code was designed and implemented with PHP Zend Framework, which save a lot of work with programming own code and ever work with MySQL database.

VII. SYSTEM PRESENTATION

Test demo version of the web application can be found at URL http://www.mediacat.sk/scyr2011/

Login of the above website are follows:

Login: demo, password: demo

Visualized model is only a demonstration. Visualized database in this article may not correspond to reality.

VIII. CONCLUSION

The database has been made in two designs, classic tabulated, but also visualized. On the final version of the database is still working and is developing. The database will be adapted for other laboratories and offices.

IX. ACKNOWLEDGMENT

REFERENCES

[1] J. Castagnetto, Professional PHP Programming, H. Rawat, S. Schumann, Ch. Scollo, D. Veliath, 2002, ISBN: 80-7226-310-2
[2] Rob Allen, *Zend Framework in Action, Nick Lo,* Steven Brown, 2008, ISBN: 1933988320
[3] Sams, *Teach Yourself MySQL in 21 days*, 2 edition, ISBN-10: 0672323923
[4] W. Jason Gilmore, *Velká kniha PHP a MySQL 5*, prepracované vydanie 2005, ISBN-10: 8086815536
[5] Nicholas C. Zachas, Jeremy McPeak, Joe Fawcelt *Profesionálne AJAX*, ISBN-10: 8086815770

WPA2 – Protocol For Higher Security In Wireless Networks

[1]Peter FANFARA, [2]Eva DANKOVÁ, [3]Marek DOMITER

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]peter.fanfara@tuke.sk, [2]eva.dankova@tuke.sk, [3]marek.domiter@tuke.sk

Abstract—Security in wireless networks is a section of informatics, which is progressing and is still being developed. Growth of computing power increases the risk of attacks in wireless networks to obtain information of any type. Increasing the security by using security protocols to protect data is necessary. Paper attention is mainly devoted to the latest protocol called WPA2, to make thorough analysis and possible solutions to correct deficiencies to increase the real-time security of Wi-Fi networks against attacks of any type.

Keywords—WEP, Wireless Network Security, WPA, WPA2.

I. INTRODUCTION

Wireless Local Area Networks (WLANs) provide the Wi-Fi communication via radio signals instead of cables used in standard wired networks. WLAN is usable over shorter distances & expands existing traditionally wired networks (LANs). WLANs are typically used in the edges of wired networks. To use WLAN is conditioned by attaching a device called the access point (AP). Communication between clients and AP in Wi-Fi network runs by using a network adapter, with implemented or attached wireless interface, which is similar to standard Ethernet adapter. Because of covering large area with radio waves & possibility to view the wireless network by anybody with Wi-Fi device in network range is security used in WLAN primary and so important. No computer network is truly secure. It's always and always will be potentially possible for eavesdroppers to view/snoop the current traffic on any wired or wireless network or add/inject undesirable traffic as well. Always are and will be some networks built & managed better than the others. [1]

The following chapters describe the security of WLANs by using the most widespread security protocols.

II. ANALYSIS OF CURRENT SITUATION

A. WLANs security

Early Wi-Fi networks worked like a Virtual Private Network (VPN) to afford security on 3rd layer of ISO/OSI, what keep IP network more vulnerable for attackers. Wireless Protected Access 2 (WPA2) is 2nd layer protocol and offers better security and protection for the network.

Wi-Fi networks add an extra level of secure complexity compared to wired networks. Whereas the wired networks send electrical signals or rays of light through cable, wireless radio signals are disseminate through the air and are naturally easier to intercept. Radio waves from access points pass across building wall and into nearby area. Security represents an important issue for WLANs. Random wireless client access must be forbidden from joining the WLAN with a sensitive content. Technologies like Wired Equivalent Privacy (WEP) increase security level on wireless networks to emulate standard wired networks.

Network engineers and experts have carefully examined security of wireless network because of the Wi-Fi communications open-air nature.

The practice of wardriving exposed the vulnerabilities of home WLANs and accelerated the pace of security technology advances in home wireless equipment. Security features like 128-bit WEP and WPA can scramble or encrypt network traffic so content cannot be easily deciphered by snoopers. [3]

Naturally every home or business must determine for the level of risk they are comfortable in taking when implementing a wireless network. The better is wireless network administered, the more secure it becomes. However, the only truly secure network is the one never built.

B. Wired Equivalent Privacy

Wired Equivalent Privacy (WEP) is a protocol that raised security of WLANs based on the IEEE 802.11 Wi-Fi standards. WEP is an ISO/OSI Data Link Layer (Layer 2) and it gives wireless networks privacy protection comparable to a standard wired network. WEP represents only one part of overall security strategy of WLAN.

WEP protocol is based on a security scheme called RC4. This cipher algorithm combine user secret key with combination of generated values. WEP implementation support 40-bit encryption – key of length 40 bits and 24 additional bits of system generated data (64 bits total). This kind of encryption solution is too weak, because it's too easy to decode. Today implementation has 128-bit encryption (104 bits key) or stronger (152-bit & 256-bit encryption). In communication over the wire, Wi-Fi equipment uses WEP keys for encryption of the data stream. Keys are stored on Wi-Fi adapter or in the Windows Registry and are not sent over the whole network. [12]

WEP was prone to attacks and poorly implemented by vendors – the main reason why it was never took off in the enterprise. The Wi-Fi Alliance because of WEP's weaknesses created WPA, a subset of the draft version 802.11i, and a few years later created WPA2, protocol providing stronger security than the first version of WPA [TABLE 1].

TABLE 1

SECURITY COMPARISON OF WEP, WPA, WPA2

	WEP	WPA	WPA2
Cipher	RC4	RC4	AES
Key Size	40 bits	128 bits encryption 64 bits authentication	128 bits
Key Life	24-bit IV	48-bit IV	48-bit IV
Packet Key	Concatenated	Mixing Function	Not Needed
Data Integrity	CRC-32	MIC	CCM
Header Integrity	None	MIC	CCM
Replay Attack	None	IV Sequence	IV Sequence
Key Management	None	EAP-based	EAP-based

C. WPA with TKIP & MIC implementation

Wireless Protected Access (WPA) is a security protocol that is used for wireless networks with port-based authentication protocol IEEE 802.1x. WPA is more secure than older protocols because it fixes the security flaws in WEP, improves authentication and encryption features and was developed by the network industry in response to the shortcomings of WEP. Despite of security progress WPA is still subject of Denial of Service (DoS) attacks.

One of the most powerful technologies in WPA is the Temporal Key Integrity Protocol (TKIP). TKIP addresses the encryption weaknesses of WEP [11]. WPA built-in authentication, that WEP doesn't offer, is another key implementation. With both technologies WPA provides roughly comparable security to VPN tunneling with WEP on the Data Link Layer (2nd Layer of ISO/OSI model), of course with easier administration and easier use.

One variation of WPA is Pre Shared Key (WPA-PSK). It is a simplified but still powerful form of WPA. This form is frequently used for home Wi-Fi networking. To use WPA-PSK, you have to sets a static key or "passphrase" same as with WEP. WPA-PSK preset, using TKIP, is automatically set to changes the keys at a time interval. This move raises the difficulty for hackers to find keys and then exploit them. TKIP uses the RC4 cipher algorithm, and it is possible to implement it in software with just a driver or firmware update. Keys are rotated frequently, and the packet counter prevents packet replay or packet re-injection attacks. WPA provides a 64-bit message integrity checking using Message Integrity Code (MIC), which is a sequence counter implemented for protection against replay attacks. MIC was designed to prevent an attacker from capturing, altering and/or resending data packets. This replaces the Cyclic Redundancy Check (CRC) that was used and implemented by the WEP standard. MIC solved problems with providing a sufficiently strong data integrity for the packet it handled in CRC with using packets integrity checker algorithm. Packets received out of order will be rejected by the AP. In a case of brute-force attacking method is network traffic automatically halted for a period of one minute and access point resets all session keys. If AP based on WPA detects more than one TKIP MIC failure within one minute, it will implement countermeasures and will rekey TKIP session key. This action is changing future keystreams and it minimizes the possible risk. [1]

III. WIRELESS PROTECTED ACCESS 2

Wi-Fi Protected Access 2 (WPA2) is a second version of WPA designed to increase the security for wireless networks. This access control technology gives Wi-Fi networks confidentiality and data integrity at the same time – two terms not previously associated with Wi-Fi. Protocol is available on all wireless network hardware since 2006. It improves the security of Wi-Fi connection by using stronger wireless encryption than older version. [8]

One of disadvantages of WPA2 is it can't alone provide enterprise security. It's possible with combining WPA2 with the IEEE 802.1x port-based authentication protocol for access control should eliminate most security worries. This solution won't protect from rogues, DoS attacks or some kind of interference but it will provide secure Wi-Fi communication.

WPA2, meanwhile, uses a Counter-Mode with CBC-MAC Protocol (CCMP) encryption method [4], which is based on Advanced Encryption Standard (AES), a stronger encryption algorithm than RC4. Specifically, WPA2 does not allow use of an algorithm called TKIP that has known security holes. However, TKIP is also strong and tends to be supported by more equipment on the market. Either is far preferable to using no encryption at all.

1) WPA2 modes [6]

Both protocols (WPA & WPA2) include two authentication modes:

I. *WPA2-Personal* generates a key of length 256-bit from a plain-text pass phrase. The PSK form the mathematical basis for the Pairwise Master Key (PMK) that's used to initiate a four-way handshake and generate the Pairwise Transient Key (PTK) between the wireless user device and access point. WPA2-Personal, like static WEP, poses challenges in key distribution and maintenance, making it a fit for small offices but not the enterprise. [3]

II. *WPA2-Enterprise*, meantime, addresses concerns regarding distributing and managing static keys, and controls access on a per-account basis by tying in to most organizations' authentication services. This authentication mode requires credentials, like user name and password, then certificate or a one-time password. Authentication runs between authentication server and station. Connection is monitored by AP or Wi-Fi controller and directs authentication packets towards the suitable authentication server.

2) WPA2 Keys

WPA2 contains different forms of security keys. *WPA2 Pre-Shared Key* (WPA2-PSK) uses 64 hexadecimal digit long keys. This method is usually applied in many home Wi-Fi networks. [9]

WPA2 also used for data encryption Group Temporal Key (GTK) & Pairwise Transient Key (PTK). Each user associated and connected to the AP with turned on WPA2 security protocol obtains a unique PTK which is used for securing unicast data traffic between AP and user. PTK is a two-way private key used for en/decryption of unicast data traffic by client of the AP. GTK is shared by all users connected to the same AP. GTK is one-way shared key which is used only for encryption of broadcasting or multicasting data traffic sent by the AP and decryption received data from AP by users.

A. Advantages & disadvantages of WPA2

Some of the biggest advantages are: Protocol gives confidentiality & data integrity, it uses stronger Wi-Fi encryption key of length at least 128-bit and it doesn't allow TKIP with its security holes in WPA implementation. Beside of advantages WPA2 has bigger system requirements, which are consequence of using stronger encryption.

B. WPA2 vulnerabilities

Every encryption system has a history of security problems especially systems used by wireless routers. WEP was cracked just a few years later then was introduced and used. Now is regarded by security expert to be entirely insecure.

WPA with TKIP was developed as kind of an interim encryption method as Wi-Fi security was evolving several years ago.

1) Cracking WPA2

Cracking networks with active WPA or WPA2 is different from cracking WEP. It's not crack in a matter of few minutes. Attacker need to capture a connection handshake from a valid user, who wants to connect to the WPA or WPA2 Wi-Fi network. After that attacker has to brute force user's connection with authority. If attacker uses Portable Penetrator for assistance which will provide enforced speed and an option to use excessively large dictionaries in many languages. [5]

There are two possible ways how to get the handshake if users are connected to WPA2 Wi-Fi network [3]:

- Get captured a handshake is possible at any moment when someone is connecting to the Wi-Fi network. If no users are connected to the wireless network, attacker has to be patient and wait for some user, who wants to connect to network in order to capture the WPA or WPA2 handshake. After handshake is capture, attacker can crack it. Otherwise, he won't be able to crack the encryption.

- If there is one or more users connected, attacker can simply choose the network and select one of the users to attack him with creating quick DoS attack and wait for disconnection of that user. When user tries to reconnect, attacker can capture the handshake. Usually user won't notice anything during the attack. When attacker has the handshake, the cracking can begin with the selected dictionary. With using a Portable Penetrator good crack will run with 1 400 keys a second – equivalent to millions within a few hours. If used dictionary doesn't contain password, attacker can select another dictionary. It is recommended to start by selecting native language dictionary or choosing a large dictionary file to increase cracking performance.

2) Wardriving and Warchalking

Wardriving is the practice of searching for WLAN signals within a geographic area. Attacker can use an automobile, GPS and strong antenna mounted onto his car to identify unsecured WLANs in near area or neighborhoods.

Some computer hackers are content to simply map any open, unsecured WLANs they find. *Warchalking* is the practice of tagging near pavement to allow others to tap in and steal bandwidth from unsecured hotspots. Wardriving is a controversial practice, but it has helped raise awareness of the importance of WLAN security. [7]

3) Hole196

The GTK, WPA2 protocol data encryption key, may be abused by a mischievous insider (must be an authorized user)

in a WPA2 wireless network. Insider can spoof the MAC address of the AP and then inject an encrypted packet using GTK which contains broadcast destination address. With GTK encrypted data is possible to launch Address Resolution Protocol (ARP) poisoning attacks.

Malicious insider uses the GTK encrypted packet in the ARP poisoning attack based on Hole196[1] to impersonate itself as the gateway to other Wi-Fi authorized users. This move provides tricking users into redirecting their data traffic to attacker via the AP. Authorized users that have received spoofed packet automatically send their data traffic (encrypted with their PTKs) to the AP and then AP forwards received data to the insider computer (encrypted with insider's PTK), where can be very easily decrypted – insider is able to bypass data privacy and he can see private data traffic from other authorized users of Wi-Fi network. The attacker doesn't require users PTKs because AP is tricked to doing all the work, what means decrypt data from victim users and forward them to the attacker by encrypting data with the attackers PTK (Figure 1).
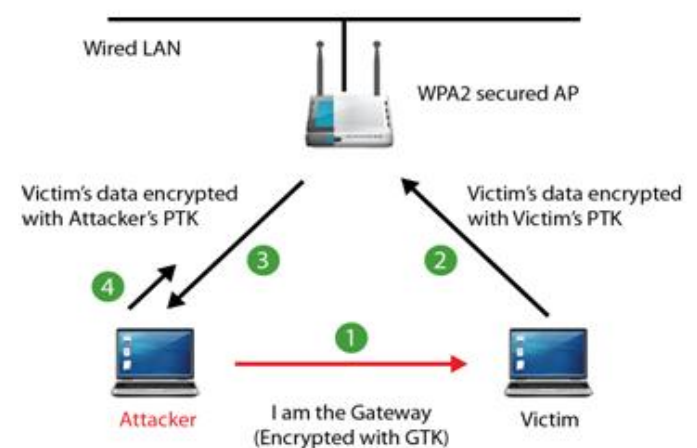

Figure 1 Hole 196 vulnerability in WPA2

1. Attacker broadcasts a fake ARP message to poison the ARP table of connected Wi-Fi users in WPA2 network, mapping IP address of the actual gateway to the MAC address of the attacker's device
2. Victim sends all its PTK encrypted data traffic to the AP with the Attacker's MAC as the destination (gateway)
3. AP forwards Victim's data to Attacker encrypting with Attacker's PTK that can decrypt private data from the Victim by using his own PTK.
4. Attacker then forwards the data traffic to the actual gateway so attack is transparent to Victim who continues communication as normal.

ARP poisoning represents attack that has possibility to launch over the Ethernet or WPA2 secured AP. It's old type of launching the attack, where AP forwards spoofed ARP messages over the Wi-Fi as well as the wired network. Today network security has evolved to the point that even network devices like switches and wired IDS/IPS may easily catch and block ARP poisoning attack on the wire. But if attacker launches an ARP poisoning attack by using spoofed GTK encrypted frames limits the footprint of the attack only to the

---

[1] The nickname 'Hole 196' refers to the page number in the IEEE 802.11 Standard where the vulnerability is buried.

air and the payload is encrypted. Then no wire-side security solution is going to catch ARP attack over WPA2 and AP will not see anything abnormal unless a Wireless Intrusion Prevention System (WIPS) is installed in place for scanning the airspace and detects strange behavior produced by the attacker over the air, this type of attack is undetectable. [10]

C. Protection against weaknesses of WPA2

There are only a few weaknesses of WPA2 security protocol. To improve level of Wi-Fi networks is appropriate to use the Portable Penetrator to audit Wi-Fi network often as is possible to guarantee the low risk of compromising or invading wireless network by hackers or crackers.

In enterprise sphere of Wi-Fi networks is security software usually included so risk of man-in-the-middle type attack is negligible. The first attack against WPA should give people a reason to dump WPA with TKIP and change from TKIP to AES encryption. Also it is possible to avoid any kind of attacks by using port security based on filtration of MAC addresses, using stronger encryption key of minimum length 256 bits or depth administering Wi-Fi network

1) Solving WPA2 Hole 196 vulnerability

Plenty access points have implemented feature called client isolation. If this feature is turned on insider is possible to send out encrypted spoofed broadcast packets to other users connected to the network, but he won't be able to receive any data traffic from a victim user, while both are connected to the same AP. This is not an ultimate solution because a variant of the ARP poisoning and man-in-the-middle attack can bypass this feature by using the Ethernet MAC address of the insider's laptop, which is connected to the LAN or by using a separate computer on wired network. Client isolation is a proprietary feature and implementation varies among WLAN vendors. It's not recommended to use client isolation because of applications running over Wi-Fi (VoIP over Wi-Fi) that require users to communicate via the AP. [7]

Vendors of WLAN APs can obviate WPA2 vulnerability Hole 196 by stopping the use of GTK. In today WLAN architectures the APs are set to no transmit broadcast traffic through the air and instead WLAN controller keeps an ARP table so GTK doesn't get used. GTK needs to be assigned by the AP, according to current WPA2 security protocol, to each user during the association process which contains from four-way handshake. WLAN AP vendors can also release patch or firmware update to AP software/devices to allocate unique GTK that will be randomly generated between all users. This patch will eliminate Hole196 vulnerability and there will be no cost related with change in conditions of decreased data load carrying capacity, unless broadcast traffic is sent by AP through the air using incomparable GTKs. In the long term, this approach of criticism the use of a shared GTK can also be adopted in the IEEE 802.11 standard.

Wi-Fi intrusion prevention system offers faster managing of emerging vulnerabilities unless suitable fixes will be implemented in the WLAN infrastructure. WIPS also offers protective layer of defense. With using WIPS is possible to detect and also focus to source of attack based on WPA2 Hole196 and then close the window of exposure unless WPA2 security protocol fix is provided in the standard by infrastructure WLAN vendors & by this way decrease the vulnerability level in Wi-Fi security protocols

IV. CONCLUSION

Implementation of new security protocol WPA2 for WLANs is difficult because implementation depends on protocol hardware support in network devices by different vendors. The 'Hole196' vulnerability exposes all Wi-Fi networks with WPA & WPA2 protocol, independent on the authentication & encryption to insider attackers. A malicious insider may exploit 'Hole196' with bypassing users data privacy and decrypt data traffic of network authorized Wi-Fi users. Also a variety of classic attack can be launched.

Enterprises can decide to use client isolation and security software situated at the endpoint, but these solutions have limited efficiency. Vendors of WLAN infrastructure can afford a proprietary solution by releases of patch or firmware to AP software/device to allocate unique GTKs to its associated users. This will eliminate 'Hole196' vulnerability without breaking the interoperability with Wi-Fi users.

Whereas vulnerability will be fixed, wireless intrusion prevention system (WIPS) may provide protection for networks 24/7/365 not only from 'Hole196', but also from other Wi-Fi threats.

The better will be wireless LAN secured the more challenging will be for attackers to find its vulnerability.

REFERENCES

[1] Practically Networked "Securing Your Wireless Network", [Online] Available: http://www.practicallynetworked.com/support/wireless_secure.htm

[2] J. Edney, W.A. Arbaugh "Real 801.11 Security: Wi-Fi Protected Access and 802.11i", Addison-Wesley 2003, ISBN: 0-321-13620-9.

[3] M.Beck, E. Tews "Practical attacks against WEP and WPA", 2008, [Online] Available: http://dl.aircrack-ng.org/breakingwepandwpa.pdf.

[4] J. Jonsson "On the Security of CTR + CBC-MAC", Retrieved 2010-05-15.

[5] G. Fleishman "Battered, but not broken: understanding the WPA crack", Ars Technica 2008, [Online] Available: http://arstechnica.com/articles/paedia/wpa-cracked.ars.

[6] F. Bulk "Learn The Basics Of WPA2 Wi-Fi Security", 2006.

[7] D. Ross "Securing IEEE 802.11 Wireless LANs", 2010.

[8] Wi-Fi Alliance "WPA2 Security Now Mandatory for Wi-Fi CERTIFIED Products", [Online] Available: http://www.wi-fi.org/pressroom_overview.php?newsid=16.

[9] J. van Rantwijk "WPA Key Calculation – From Passphrase To Hexadecimal Key", Retreived 2009-01-16, [Online] Available: http://www.xs4all.nl/~rjoris/wpapsk.html.

[10] C. Nachreiner "Anatomy of an ARP Poisoning Attack", [Online] Available: http://www.watchguard.com/infocenter/editorial/135324.asp

[11] M.F. Halvorsen, O.Haugen,M.Eian;S.F.Mjølsnes "An Improved Attack on TKIP" September 2009, Lecture Notes in Computer Science, 2009, Volume 5838/2009, pp. 120-132, doi:10.1007/978-3-642-04766-4_9.

[12] Wi-Fi "Wi-Fi Alliance Announces Standards-Based Security Solution to Replace WEP", 2002, [Online] Available: http://wi-fi.org/pressroom_overview.php?newsid=31.

[13] L. Vokorokos, N. Ádám, A. Baláž, J. Perháč "High-Performance Intrusion Detection System for Security Threats Identification in Computer Networks", Elfa s.r.o. 2009, ISBN: 978-80-8086-131-5.

Author's index

11th Scientific Conference of Young Researchers
of Faculty of Electrical Engineering and Informatics
Technical University of Košice

Proceedings from Conference