



12th Scientific Conference of Young Researchers

May 15th, 2012
Herľany, Slovakia

Proceedings from Conference

Faculty of Electrical Engineering and Informatics
Technical University of Košice



Sponsors



**12th Scientific Conference of Young Researchers
of Faculty of Electrical Engineering and Informatics
Technical University of Košice**

Proceedings from Conference

Published: Faculty of Electrical Engineering and Informatics
Technical University of Košice
I. Edition, 440 pages

Available: http://web.tuke.sk/scyr/data/templates/Proceedings_2012.pdf

Editors: Prof. Ing. Alena Pietriková, CSc.
Ing. Dominik Demeter
Ing. Milan Nosál

ISBN 978-80-553-0943-9

**Program Committee of 12th 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.
Assoc. Prof. Ing. František Ďurovský, PhD.
Assoc. Prof. Ing. Ján Gamec, CSc.
Assoc. Prof. Ing. Ján Genčí, PhD.
Assoc. Prof. Ing. Jozef Juhár, PhD.
prof. Ing. Iraidia Kolcunová, PhD.
Assoc. Prof. RNDr. Ján Ziman, CSc.
prof. Ing. Ján Paralič, PhD.
Ing. Marek Paralič, PhD.
Assoc. Prof. Ing. Daniela Perduková, PhD.
prof. Ing. Alena Pietriková, CSc.
Assoc. Prof. Ing. Jaroslav Porubán, PhD.
Assoc. Prof. Ing. Branislav Sobota, PhD.

**Organization Committee of 12th Scientific Conference of Young Researchers of
Faculty of Electrical Engineering and Informatics Technical University of Košice**

Members: Ing. Martin Bačko
Ing. Mišél Batmend
Ing. Dominik Demeter
Mgr. Peter Duranka
Ing. Ivan Halupka
Ing. Pavol Hocko
Ing. Jozef Lipták
Ing. Slávka Jadlovská
Ing. Milan Nosál
Ing. Adela Tušanová
Ing. Matúš Valo

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 multicultural atmosphere. The Eleventh Scientific Conference of Young Researchers (SCYR 2012), conference of Graduates and Young researchers, was held on 15th May 2012. 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. 110 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.

12th Scientific Conference of Young Researchers at Faculty of Electrical Engineering and Informatics Technical University of Košice (SCYR 2012) 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 approximately 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.

Liberios VOKOROKOS
dean of FEI TUKE

May 15th 2012, Herlany

Contents

Informatics & Telecommunications – Oral form

Peter VISZLAY

Eigenvalue–Sensitive Partial Subspace Training for Large Speech Corpora 15

Michaela BAČÍKOVÁ

Domain Analysis and Feature Modeling of GUIs 19

Vladimír CIPOV, Marek COPÁK

Anchor-free positioning using ToA node distance estimation and LoS/NLoS detection 23

Mária VIRČÍKOVÁ, Peter SMOLÁR

Current Trends in Human–Robot Interaction: Towards Collaborative & Friendly Machines 27

Emília PIETRIKOVÁ, Sergej CHODAREV

Experiments on Language Patterns for New Approach in Formation of Programming Languages
31

Štefan JAJČIŠIN

Explicit and Online Predictive Control of Nonlinear Hydraulic System 34

Slávka JADLOVSKÁ

Swing-up and Stabilizing Control of Classical and Rotary Inverted Pendulum Systems 38

Milan NOSÁL

Tool Support for Task-driven Case Study Design and Evolution 42

Informatics & Telecommunications – Poster form

Eva VOZÁRIKOVÁ

Acoustic Event Detection using MFCC and Genetic Algorithm 48

Gabriel LUKÁČ

Crawling dynamic online networks 52

Anton MOLČAN

Application of 3D sensors in the design of automatic unloading mechanisms 56

Sergej CHODAREV

Stratified Framework for DSL Development 60

Alexandra LUKÁČOVÁ

A review of data mining applications in manufacturing 64

Emília PIETRIKOVÁ, Ivan HALUPKA

Analysis of Abstraction and Structural Complexity Within Domain of Program Samples .. 68

Matej ČOPÍK, Ján ILKOVIČ

Analysis of manufacturing systems using Petri nets 72

Marek NOVÁK, Martin KAPA

Anomalies detection in user's motion patterns in a smart-home environment 76

Jozef VAVREK

Audio Content Classification using SVM Binary Decision Trees 80

Ivan HALUPKA, Emília PIETRIKOVÁ

Automated Refactoring algorithm – AURA 84

Viliam ROČKAI, Marián MACH

Automatic concept identification in slovak texts 88

Dominik LAKATOŠ

Basis for Model Oriented Language Composition 92

Marián JENČIK

Complex program system from ABT view 96

Martin REPKA

Component Identification in Company Networks 99

Adrián PEKÁR, Marek DUFALA

Contribution to Real-Time Network Traffic Evaluation 103

Peter IVANČÁK, František HROZEK

Creation of Depth Map for 3D Display 107

Adela TUŠANOVÁ

Decision-making framework for adoption of cloud computing 110

Anton BALÁŽ

Denial of Service Intrusion Detection System 113

Lukáš LACIŇÁK, Roman MIHAL

Design and Development of Application for Supervisory Control 117

Vojtech RINÍK, Ivan KLIMEK <i>Edukit: A Real-Time Web-Based Collaboration Tool for Educational Purposes</i>	121
Eva DANKOVÁ, Adrián PEKÁR <i>ETL System Proposal</i>	124
Anton BALÁŽ <i>Extract – Transform – Load Data Warehouse System</i>	129
Denis DUPÁK, Radovan BLIČA <i>Filter Bank Multicarrier Transmission System</i>	133
Vladimír SERBÁK <i>Fuzzy control application for a class of nonlinear systems</i>	136
Peter ADAMONDY, Peter PAPCUN <i>Gesture recognition using Static Bayesian Tree algorithm</i>	140
Michal ENNERT, Peter IVANČÁK <i>GPGPU: The Increase of the Computing Power for Data Encryption</i>	143
Mišel BATMEND, Martin MIKULÁK <i>Hamiltonian shortest path approximation optimizing a bitmap engraver tool path</i>	148
Pavol MACKO <i>How to apply recursion and corecursion in mathematical theory of programming</i>	151
Ján VALISKA, Ondrej KOVÁČ <i>Channel parameters simulation using NS-2 and MyEvalvid-NT environments</i>	154
Vladimír GAŠPAR, Tomáš KAROL, Cecília HAVRILOVÁ <i>Ideal MPM-20 jet engine running process with zero degrees of freedom</i>	158
Marek ČAJKOVSKÝ, Ivan KLIMEK, Michal ENNERT <i>Intrusion Detection Systems – Overview, Timeline and Techniques</i>	162
Ivan KLIMEK, Marián KELTIKA <i>Jitter Utilizing Congestion Estimation (JUICE)</i>	166
Ján ŠTOFA, Dana PALOVÁ <i>Management of enterprise security</i>	169
Martin VARGA <i>Markerless augmented reality using SURF method</i>	173
Ján ILKOVIČ, Peter PAPCUN <i>Material flow in Flexible Assembly Company</i>	177
Martin PAĽA, Ladislav MIŽENKO <i>Mobile Robot Navigation using Adaptive Fuzzy Cognitive Maps</i>	181
Dávid CYMBALÁK <i>Multimedia solution for live streaming of various types of educational content</i>	185
Peter ŠUSTER <i>Nonlinear Control Design of Robot Arm</i>	188
Peter SMOLÁR, Mária VIRČÍKOVÁ <i>Object Categorization using system Multi-Agent Serving System and robot Nao</i>	192

Ivan PETŤKO	
<i>On decomposition of High Level Petri Nets</i>	196
Matúš VALO	
<i>On Kronecker product of graphs</i>	200
Milan NOSÁĽ	
<i>Overview of Literate and Elucidative Programming</i>	204
Patrik GALLO	
<i>Parallelizing computations over elliptic curve arithmetic in affine and Jacobian projective coordinates</i>	208
Peter PAPCUN, Matej ČOPÍK	
<i>Remote control of Mitsubishi industrial robot</i>	212
Martina LAĽOVÁ, Lukáš LACIŇÁK	
<i>SCADA/HMI System in Action Graphs</i>	216
Dávid CYMBALÁK	
<i>Solution for receiving live broadcast of lecture with interaction between teacher and students</i>	219
Daniel GONTKOVIČ	
<i>State Control Design for Systems with Time Delay</i>	222
Miloš PAVLÍK, Rastislav HOŠÁK	
<i>Testing the vSphere Storage Appliance in Virtual Environment</i>	225
Kristián ŠESTÁK	
<i>Three-Dimensional Software Visualization and UML</i>	228
Peter FANFARA, Adrián PEKÁR	
<i>Usage of Hybrid Honeypot as an Intrusion Detection System Mechanism</i>	233

Electrical & Electronics Engineering – Oral form

Marek PÁSTOR

Grid-tied Multilevel Inverter for Photovoltaic System 239

Martin LIPTAJ, Tomáš HRÁŠOK

Design of the 5-bit Flash A/D Converter for UWB Applications 243

Matúš OCILKA, Oleksii KRAVETZ, Oleksander SHUTKA

Calculation of Inductance of Double-Layer Planar Spiral Coil Using COMSOL Multiphysics
247

Martin SEKERÁK, Marek GODLA, Jozef LIPTÁK

DAC Testing by using a precise Reference DC Voltage and Dithering Signal 250

Martin GERMAN-SOBEK, Jozef KIRÁLY

Dielectric Spectroscopy of XLPE Cables 255

Jozef KIRÁLY, Martin GERMAN-SOBEK

Influence of concentration of magnetic nanoparticles in transformer oil based ferrofluid from view of dielectric spectroscopy 259

Daniel URDZÍK

Simulation of shadowing effect for UWB through wall surveillance in Comsol Multiphysics environment 263

Peter DURANKA, Magdaléna UHRÍNOVÁ

Study of isotactic polypropylenes by magic angle spinning ^1H nuclear magnetic resonance 267

Electrical & Electronics Engineering – Poster form

Karol KYSLAN, Matúš HRIC

A Novel Approach to Dynamic Emulation of Mechanical Loads Based on Industrial Converters' Features 271

Ján PERDULAK, Oksana KHREBTOVA, Oleksii LESHCHUK

A Novel Multiphase Boost Converter with High Efficiency of Energy Conversion 275

Jozef DUDIÁK, Pavol HOCKO

Modeling support services by using Modes 279

Michal KRAVČÍK

Printing Performance of Solder Paste Dependent on Temperature and Density 282

Dmytro MAMCHUR, Andrii KALINOV, Ján MOLNÁR

A comparison the effectiveness of induction motor's diagnostic methods based on spectra analysis of current and instantaneous power signals 286

Jozef LIPTÁK, Marek GODLA, Daniel FÁBRI

ADC test stand with exponential excitation signal using LabVIEW 291

Ondrej KOVÁČ, Kornel RUMAN, Ján VALISKA

Application binary state arithmetic encoding in JPEG 295

Pavol MIŠENČÍK, Tomáš HARASTHY

Availability of Hybrid FSO/RF System 299

Matúš HRIC, Karol KYSLAN

Basic Nonlinear Model of Precision Gearbox 302

Tomáš HARASTHY, Pavol MIŠENČÍK

Cambridge Correlator in Traffic Sign Recognition System 306

Tibor BALOGH

Comparison of Motor Dynamics of DC and BLDC Motors 309

Peter NGUYEN, Marek VACEK

Comparison of software for modeling and simulation of robotic arms in the university environment 313

Michal JURČIŠIN, Pavol CABÚK

Constructional solution of sensors for measuring of intra-abdominal pressure 316

Kornel RUMAN, Ondrej KOVÁČ, Igor VEHEC

Design of Microstrip Hairpin Band Pass Filter for UWB Application 319

Martin LIPTAJ, Matej ŽIGA

Design of the I-Q Demodulator for UWB Applications 322

Viktória ŠUHAJOVÁ

Domain Walls Contribution to the Magneto-Impedance Effect in Amorphous Magnetic Wire
326

Marek PAVLÍK, Ján ZBOJOVSKÝ

Effect of electromagnetic field 329

Marek GODLA, Daniel FABRI, Jozef LIPTÁK

<i>Effective Measurement Stand for Teaching Basic Electronic Circuits</i>	333
Nataliia ISTOMINA, Matúš OCILKA, Ján MOLNÁR	
<i>Electromagnetic processes of switched reluctance motor with variable phase inductance</i> ...	337
Martin BAČKO, Oksana BRATASH, Vita OGAR	
<i>Energy output of photovoltaics and wind energy in Košice region during winter months</i> ..	340
Tomáš KOŠICKÝ	
<i>Frequency regulation with energy storage</i>	344
Michal KALAVSKÝ	
<i>Harmonic Potential Field Method for Path Planning of Mobile Robot</i>	348
Tomáš BÉREŠ, Michal KALAVSKÝ	
<i>Improved experimental model of bi-directional DC/DC Converter</i>	352
Zsolt ČONKA, Stanislav KUŠNÍR	
<i>Improving transient stability with utilizing FACTS device</i>	355
Vladimír KRIŠTOF, Stanislav KUŠNÍR	
<i>Influence of PST transformers on short-circuit condition of the electric power system</i> ...	359
Marián HALAJ	
<i>Influence of the automatic transfer switch (ATS) on the power supply reliability</i>	362
Gabriela VASZIOVÁ, Lukáš GLOD	
<i>Interpretation of regular force measurements in fluids</i>	365
Dmytro BORDUG, Martin BAČKO, Tibor VINCE	
<i>Laboratory complex for researching of the cooling systems for electric machines</i>	369
Magdaléna UHRÍNOVÁ, Peter DURANKA	
<i>Laboratory frame ^1H spin-lattice relaxation study of isotactic polypropylenes</i>	372
Michal KALAVSKÝ	
<i>Mathematical Analysis of Harmonic Potential Field Method</i>	375
Matúš TATARKO	
<i>Measuring of Availability and Reliability of FSO links from Measured Weather parameters</i>	379
Rastislav KOKOŠKA	
<i>MHP application in interactive TV</i>	382
Marcel BODOR, Milan LACKO	
<i>Novel soft switching DC/DC converter</i>	388
Dominik DEMETER	
<i>Overview of the selected technologies for creating the virtual laboratory for the assembling technologies in electronics</i>	392
Marián HRINKO, Marek PAVLÍK	
<i>Partial discharges pattern statistical analysis</i>	396
František RAKOCI	
<i>Road Signs Recognition and Inventory System using Optical Correlator</i>	400
Peter NGUYEN, Marek VACEK	
<i>Sensorless vector control of induction motor with neural speed estimator</i>	403

Ján ZBOJOVSKÝ, Marek PAVLÍK <i>Sources, use and measurement of electromagnetic field</i>	407
Iurii ZACHEPA, Ján PERDULAK, Tibor VINCE <i>Specificity of DC consumers in independent power supplies based on an asynchronous generator</i> 411	
Matúš KATIN, Roman JAKUBČÁK <i>Swing up of conductor depending on weight of ice and span size</i>	414
Dmytro MAMCHUR, Andrii KALINOV, Tibor VINCE <i>The development of experimental equipment for investigation the induction motors diagnostic system based on the electrical signals analysis</i>	417
Lukáš GLOD, Gabriela VASZIOVÁ <i>The diffusion coefficient for nonlinear Brownian motion and the electric circuits</i>	421
Yuriy ROMASHIHIN, Martin BAČKO, Ján MOLNÁR <i>The energy method for identification the electromagnetic parameters of induction motors</i>	424
Pavol HOCKO, Matúš NOVÁK <i>Transient stability of generator in simple power system</i>	427
Matúš NOVÁK, Pavol HOCKO <i>Transient stability of power system</i>	431
Roman JAKUBČÁK, Matúš KATIN <i>Using the TCSC for power flow control and reduce power losses in power system</i>	435
Author's Index	438

Section:
Informatics & Telecommunications
Oral form

Eigenvalue–Sensitive Partial Subspace Training for Large Speech Corpora

Peter VISZLAY (3rd year)

Supervisor: Jozef JUHÁR

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

peter.viszlay@tuke.sk, jozef.juhar@tuke.sk

Abstract—Principal Component Analysis (PCA) is a well-known feature extraction and dimensionality reduction method applied also in speech recognition. One of the integral parts of PCA is the covariance matrix computing from the training set represented by data matrix. In case of large speech corpus this operation may be memory intensive. In this paper we propose a partial-data training procedure for transformation matrix of PCA based on eigenvalue–sensitive feature selection. Only the selected part of the whole training data is then used to PCA learning. The selective process is based also on PCA without projection. The experiments show that from the limited data subset can be PCA trained with comparative or even better performance.

Keywords—covariance matrix, eigenvalue, feature selection, transformation.

I. INTRODUCTION

Linear feature transformations are well-used techniques in high-dimensional data processing such as face and automatic speech recognition (ASR). The most popular transformations in automatic speech recognition are Principal Component Analysis (PCA) [1], [2], [3] and Linear Discriminant Analysis (LDA) [4]. Our speech recognition research group tends to follow the modern trends in ASR. Therefore, we are interested in research and application of linear transformations in our speech recognition system.

It is known that one integral part of PCA is the covariance matrix computing from the training set. In case of relatively small training corpus there is no problem to compute the covariance matrix. But, in case of large corpus (thousands of recordings) and high-dimensional data there may occur a problem related to processing time (\approx several hours) consumption and memory requirements (\approx 20GB).

In order to solve the above mentioned problems we have built upon our previous work [5], [6] and we proposed a procedure, which allow to train the PCA from a limited amount of the training data. In other words, PCA can be learned from a training subset, while the performance is maintained, or even improved. We called this procedure as *Partial-data trained PCA*. Detailed description of the proposed method is given in Section 1. This procedure is applied to LMFE (Logarithmic Mel–Filter Energies) feature vectors. The performance of the method is evaluated on Slovak speech corpus in phoneme–based continuous speech recognition task.

This paper is organized as follows. The next section describes the PCA and also highlights the proposed method. Section III presents the experimental setup and the Section IV presents our results.

II. PRINCIPAL COMPONENT ANALYSIS IN ASR

Principal component analysis (PCA) [2] is a linear feature transformation and dimensionality reduction method, which maps the n -dimensional input possibly correlated data to K -dimensional ($K < n$) linearly uncorrelated variables (mutually independent principal components) with respect to the variability. PCA converts the data by a linear orthogonal transformation using the first few principal components, which usually represent about 80% of the overall variance. The principal component basis minimizes the mean square error of approximating the data. This linear basis can be obtained by application of an eigendecomposition to the global covariance matrix estimated from the original data.

A. Mathematical description

The characteristic mathematical stages of PCA can be briefly described as follows [2], [7]. Firstly suppose that the training data are represented by M n -dimensional feature vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$. One of the integral parts of PCA is the centering of all vectors (subtracting the mean) as:

$$\Phi_i = \mathbf{x}_i - \bar{\mathbf{x}}, \quad i \in \langle 1; M \rangle, \quad (1)$$

where

$$\bar{\mathbf{x}} = \frac{1}{M} \sum_{i=1}^M \mathbf{x}_i \quad (2)$$

is the training mean vector. From the centered vectors Φ_i the centered data matrix with dimension $n \times M$ is created as:

$$A = [\Phi_1 \Phi_2 \dots \Phi_M]. \quad (3)$$

To represent the variance of the data across different dimensions, the global covariance matrix is computed as:

$$C = \frac{1}{M-1} \sum_{i=1}^M \Phi_i \Phi_i^T = \frac{1}{M-1} \sum_{i=1}^M (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T. \quad (4)$$

An eigendecomposition (5) is applied to the covariance matrix in order to obtain its eigenvectors (spectral basis) $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ and their corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ as follows:

$$C\mathbf{u}_i = \lambda_i \mathbf{u}_i, \quad i \in \langle 1; n \rangle. \quad (5)$$

The principal components are represented by the eigenvectors and the most significant ones are determined by K leading eigenvectors resulting from the decomposition. The

dimensionality reduction step is performed by keeping only the eigenvectors corresponding to the K largest eigenvalues ($K < n$). These eigenvectors form the transformation matrix U_K with dimension $n \times K$:

$$U_K = [\mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_K], \quad (6)$$

while $\lambda_1 > \lambda_2 > \dots > \lambda_n$. Finally, the linear transformation $\mathbf{R}_n \rightarrow \mathbf{R}_K$ is computed as:

$$\mathbf{y}_i = U_K^T \Phi_i = U_K^T (\mathbf{x}_i - \bar{\mathbf{x}}), \quad i \in \langle 1; M \rangle, \quad (7)$$

where \mathbf{y}_i represents the transformed feature vector. The value of K can be chosen as needed or according to the following comparative criterion:

$$\frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^n \lambda_i} > T, \quad (8)$$

where the threshold $T \in \langle 0.9; 0.95 \rangle$. T represents the part of the global variance of the original data preserved in the new feature space. Since

$$\sum_{i=1}^n \lambda_i = \text{trace}(C), \quad (9)$$

the comparative criterion can be rewritten as:

$$\frac{\sum_{i=1}^K \lambda_i}{\text{trace}(C)} > T. \quad (10)$$

B. Classical PCA in ASR

In this section we describe the classical PCA trained from the whole amount of training data (for more details about training data, see Section III-A). Each parametrized speech signal in the corpus is represented by a separate LMFE matrix $X^{(i)}$, $i \in \langle 1; N \rangle$ with dimension $26 \times n_i$ (see Section III-B), where n_i represents the number of frames in i -th recording. In our case $N = 36917$ (number of training utterances).

At the first stage, the initial data preparation is performed, which requires the mathematical computations described by (1-3). The global covariance matrix is computed according to (4) and then decomposed to a set of eigenvector-eigenvalue pairs. According to the K largest eigenvalues the corresponding eigenvectors were chosen. These ones formed the transformation matrix U_K (6), which was used to transform the train and test corpus into PCA feature space. Note that the final dimension (K) of the feature vectors after PCA transformation was chosen independently from the criterion formula (8). Detailed reasons are given in Sections III-B. However, for interest, the determined optimal dimensions for different PCA configurations computed by (8) are listed in Section IV-B.

C. Partial-data trained PCA

In case of relatively small training corpus there is no problem to compute the covariance matrix. But, in case of large corpora (thousands of recordings) and high-dimensional data there may occur a problem related to processing time consumption and memory requirements (note that PCA requires

to reside the whole data matrix in the main memory). We found that for PCA learning is not necessary to use the whole training data but it may be sufficient to use a part of them. In other words, PCA can be trained from limited (reduced) amount of training data, while the performance is maintained, or even improved.

Partial-data PCA training can be viewed as a kind of feature selection process. The main idea is to select the statistically significant data (feature vectors) from the whole amount of training data. There are two major processing stages. The first stage is the data selection based on "fast" PCA separately applied to all training feature vectors. Suitable vectors are concatenated into one train matrix, which is treated as the input for the main PCA. The second stage is the main PCA (see Section II-A).

Suppose now that apply the same conditions as in Section II-A. Then the selection process based on PCA (without projecting phase) can be described as follows. Each 26-dimensional LMFE feature vector \mathbf{x}_i , $i \in \langle 1; M \rangle$ (see Section III-B) is reshaped to its matrix version X_i , $i \in \langle 1; M \rangle$ with dimension 2×13 . After mean subtraction, the covariance matrix is computed as:

$$C_i = \frac{1}{k-1} X_i X_i^T, \quad i \in \langle 1; M \rangle; k = 13. \quad (11)$$

In the next step, the eigendecomposition is performed on each covariance matrix C_i , which results in a set of eigenvectors $\mathbf{w}_1, \mathbf{w}_2$ and eigenvalues α_1, α_2 :

$$C_i \mathbf{w}_{ij} = \alpha_{ij} \mathbf{w}_{ij}, \quad i \in \langle 1; M \rangle, j \in \langle 1; 2 \rangle, \quad (12)$$

where

$$W_i = [\mathbf{w}_{i1} \mathbf{w}_{i2}]. \quad (13)$$

Note that the parameters $\mathbf{w}_{i1}, \mathbf{w}_{i2}$ and α_{i1}, α_{i2} at each iteration i are updated with new parameters resulting from a new eigendecomposition.

For PCA-based selection process the eigenvectors $\mathbf{w}_{i1}, \mathbf{w}_{i2}$ are not used. On the other hand, the eigenvalues α_{i1}, α_{i2} are very important because the selective criterion is based exactly on them. Using the eigenvalues, the percentage proportion P_i is computed as:

$$P_i = \frac{\alpha_{i1}}{\sum_{j=1}^2 \alpha_{ij}} = \frac{\alpha_{i1}}{\alpha_{i1} + \alpha_{i2}} = \frac{\alpha_{i1}}{\text{trace}(C_i)}, \quad (14)$$

which determines the percentage of variance explained by the first eigenvalue in the whole eigenspectrum. Further, it is necessary to choose the threshold T . It can be chosen from two different intervals. The first one is defined as $T_1 \in (50; \approx 65)$ and the second one as $T_2 \in (\approx 90; 99.9)$. Then the selective criterion can be based on the following logical expressions:

$$P_i \leq T_1 \quad (15)$$

for the first interval, or

$$P_i \geq T_2 \quad (16)$$

for the second interval. If the evaluation of the expression yields a logical true then the current feature vector is classified as statistically significant for PCA training. This vector is stored and the selection process continues for the next feature vector. In this way, the whole training corpus is processed. From the selected vectors a training matrix is composed,

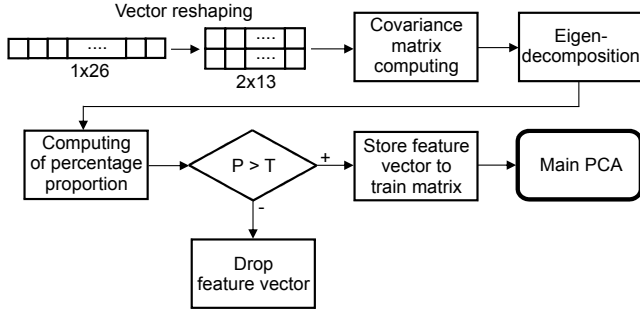


Fig. 1. Block diagram of the partial-data PCA training procedure

which is treated as the input for the main PCA described in Section II-A. As was mentioned in Section II-A, there are M training vectors in the corpus. If the selected subset contains M' vectors ($M' \ll M$, $M = 18\,981\,834$) then (3) can be modified as:

$$A' = [\phi_1 \phi_2 \dots \phi_{M'}], \quad (17)$$

where ϕ_i is the mean subtracted feature vector in the new train matrix. The next mathematical computations are identical with (4–7). The partial-data training procedure is illustrated on the Fig. 1. The new train matrix is a radically-reduced, more relevant representation of the training corpus. It has a nearly homoscedastic variance structure because it contains only those feature vectors, which have almost the same variance distribution. Feature vectors selected from the interval represented by threshold T_1 can be characterized as data clusters, which have very small variance distribution explained by the first eigenvalue among the direction of the corresponding first eigenvector. On the other hand, the feature vectors from the interval represented by threshold T_2 are clusters, which have large variance distribution among the first eigenvector. In both cases, the largeness of the variance is determined by the first eigenvalue. The size of the selected partial subset depends on the value of T_1 and T_2 . The size of partial subset can be expressed in percentage amount as:

$$subset_size = \frac{M'}{M} \times 100. \quad (18)$$

We found that a practical importance has a ratio, when

$$\frac{M'}{M} \in \langle 0.001; 0.15 \rangle, \quad (19)$$

so the selected subset contains maximally 15% of data of the whole training data amount. For example, there are approximately 19 million training vectors in our corpus. According to (19) it is sufficient to extract ≈ 19000 vectors (and more) for partial-data training. The time consumption and memory costs of the covariance matrix computation of the reduced data set are much smaller (units of megabytes) than the costs of the covariance matrix computation in case of the whole corpus.

III. SPEECH CORPUS AND EXPERIMENTAL CONDITIONS

A. Speech corpus

The speech corpus [8] contains approximately 100 hours spontaneous parliamentary speech recorded from 120 speakers (90% of men). For acoustic modeling 36917 training utterances were exactly used. For testing purposes another 884 utterances were used.

TABLE I
USED PARAMETERS FOR PARTIAL-DATA PCA MODELS

Approx. DB size	0.1%	1%	5%	10%	15%
Num. of vectors M'	22229	187248	947804	1936764	2842838
Threshold T_1	51.40	54.05	59.10	63.00	65.75
Opt. dimension d	$d = 8$	$d = 8$	$d = 8$	$d = 8$	$d = 9$
Approx. DB size	0.1%	1%	5%	10%	15%
Num. of vectors M'	23547	206624	962434	1899584	2849321
Threshold T_2	98.0	96.1	93.2	91.0	89.2
Opt. dimension d	$d = 5$	$d = 6$	$d = 7$	$d = 8$	$d = 8$

B. Speech preprocessing

The speech signal was preemphasized and windowed using Hamming window. The window size was set to 25ms and the step size was 10ms. Fast Fourier transform was applied to the windowed segments. Mel-filterbank analysis with 26 channels was followed by logarithm application to the linear filter outputs. The 26-dimensional LMFE features were used for PCA-based processing. After PCA, 13 coeffs. were retained and expanded by delta and acceleration coeffs.

C. Acoustic modeling and evaluation

The acoustic modeling by using HTK Toolkit [9] was performed. The recognition system used context independent monophones modeled using a three-state left-to-right HMMs. The number of Gaussian mixtures per state was a power of 2, starting from 1 to 256. The phone segmentation of 45 phones was obtained from embedded training and automatic phone alignment. The number of trained monophone models corresponded to the number of phonemes. During the test it was used a word lattice created from a bigram language model, which from the test set was built. The vocabulary size was 125000 words. Notice that the accuracies in the evaluation process were computed as the ratio of number of all word matches to number of reference words.

IV. EXPERIMENTAL EVALUATION AND RESULTS

A. Experiments based on full-data trained PCA

In this section, we experimentally evaluate the performance of the full-data PCA method by using the whole amount of training data. As it was mentioned in Section 1, PCA requires to allocate the whole data matrix in the memory. In addition, the covariance matrix is computed from this data matrix, which may be a computationally very difficult operation. In order to compare the partial-data trained models with the full-data trained model it was necessary to do the above mentioned computation. The full-data trained PCA was performed on a Linux machine with 32GB memory. The training data were loaded in the memory sequentially by data blocks and then concatenated to one data matrix (3). From this matrix the covariance matrix according to (4) was computed. Then the integral parts of PCA according to (5–7) were performed.

In the next step, acoustic modeling based on the PCA transformed features was done. The evaluation results of the full-data trained PCA-based model are listed in the Table II.

B. Experiments based on Partial-data PCA

The selective process for the feature vectors is performed according to Fig. 1 and it is M -times repeated. Overall,

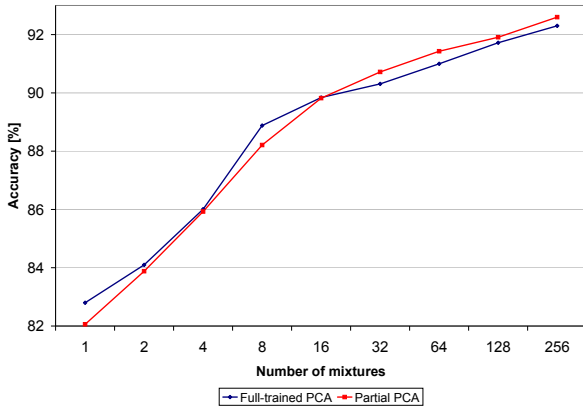


Fig. 2. Accuracy levels for full-data trained PCA and partial-data PCA.

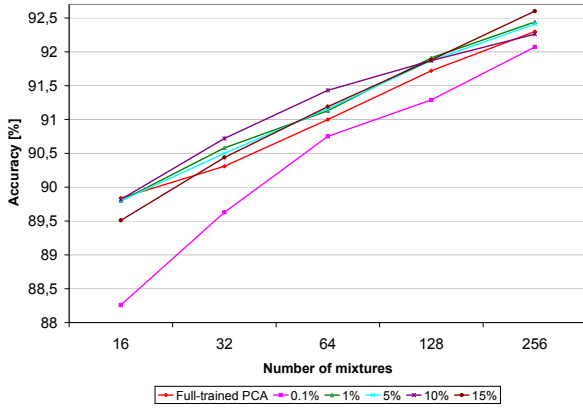
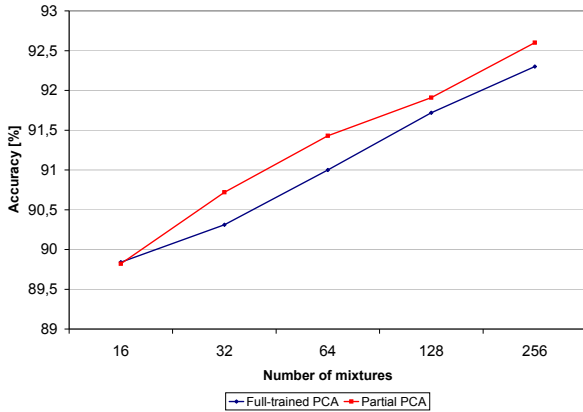
Fig. 3. Accuracies of 5 partial-data trained PCA models with threshold T_2 (16 – 256 mixtures).

Fig. 4. The highest accuracy levels for 16 – 256 mixtures.

10 partial-data trained models were learned, 5 models for selection based on threshold T_1 and 5 models for threshold T_2 . The detailed parameters of these models are listed in the Table I. According to (19), 5 subset models (0.1%, 1%, 5%, 10% and 15%) were composed. One of the output parameters of the partial-data PCA is the optimal dimension d determined by (8). It represents the number of principal components, which could be used to transform the input data with retaining 95% of global variance. Note that T_1 and T_2 were determined on experimental basis. The results of the partial models are listed in the Table II. The table contains only the highest accuracies chosen from all models. The advantage of the partial-data

TABLE II
ACCURACY LEVELS FOR FULL AND PARTIAL-DATA TRAINED PCA

Mixtures	Acc. of full PCA	Acc. of partial PCA	Difference	Threshold	Part of DB
1	82.80%	82.06%	-0.74%	$T_2 = 89.2$	15%
2	84.10%	83.88%	-0.22%	$T_2 = 89.2$	15%
4	86.01%	85.93%	-0.08%	$T_1 = 63.0$	10%
8	88.88%	88.21%	-0.67%	$T_2 = 89.2$	15%
16	89.84%	89.82%	-0.02%	$T_1 = 91.0$	10%
32	90.31%	90.72%	+0.41%	$T_1 = 91.0$	10%
64	91.00%	91.43%	+0.43%	$T_1 = 91.0$	10%
128	91.72%	91.91%	+0.19%	$T_2 = 96.1$	1%
256	92.30%	92.60%	+0.30%	$T_2 = 89.2$	15%

training is that it does not require the loading of the whole data matrix in the main memory.

C. Comparison of results

From the Table II we can conclude that the performance of the full-data trained PCA can be outperformed by the partial-data PCA. The highest accuracies for all mixtures are viewed in the Fig. 2. These cases are characteristic for higher number of mixtures, especially for mixtures 32 – 256 (see Fig. 3 and Fig. 4). The highest improvement is 0.43% abs. for 64 mixtures. For mixtures 1 – 16 the performances were not improved. We argue that the selected data were not sufficient for these mixtures. The case for 16 mixtures is characteristic one because the performances for both models are almost the same. It should be noted that for practical purposes are usually used the acoustic models with 16 – 256 mixtures.

ACKNOWLEDGMENT

The research presented in this paper was supported by the Ministry of Education under the research project VEGA 1/0386/12 (30%) and Research and Development Operational Program funded by the ERDF under the projects ITMS-26220120030 (35%) and ITMS-26220220155 (35%).

REFERENCES

- [1] H. Abbasian, B. A. Nasersharif, and A. Akbari, "Class-dependent PCA optimization using genetic programming for robust MFCC extraction," in *Proc. of the 3rd Conf. on Information and Knowledge Technology, IKT'07*, Sep. 2007.
- [2] I. T. Jolliffe, *Principal Component Analysis*. New York, USA: Springer-Verlag, 1986.
- [3] X. Wang and D. O'Shaughnessy, "Improving the efficiency of automatic speech recognition by feature transformation and dimensionality reduction," in *Proc. of the 8th European Conf. on Speech Communication and Technology, EUROSPEECH – INTERSPEECH*, Geneva, Switzerland, Sep. 2003, pp. 1025–1028.
- [4] R. Haeb-Umbach and H. Ney, "Linear discriminant analysis for improved large vocabulary continuous speech recognition," in *Proc. of the IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, ICASSP'92*, San Francisco, CA, 1992, pp. 13–16.
- [5] P. Vizslay, M. Pleva, J. Juhár, and J. Staš, "PCA-based acoustic model with logarithmic mel-filterbank features," in *Proc. of the 12th Intl. Conf. on Research in Telecommunication Technologies, RTT'10*, Velké Losiny, Czech Republic, Sep. 2010.
- [6] P. Vizslay and J. Juhár, "Feature Selection for Partial Training of Transformation Matrix in PCA," in *Proc. of the 13th Intl. Conf. on Research in Telecommunication Technologies, RTT'11*, Czech Republic, 2011.
- [7] G. Bebis, *Principal Components Analysis*, Department of Computer Science, University of Nevada, Reno, 2003.
- [8] S. Darjaa, M. Cernák, Š. Beňuš, M. Rusko, R. Sabo, and M. Trnka, *Rule-based triphone mapping for acoustic modeling in automatic speech recognition*, ser. LNCS, 2011, vol. 6836 LNAI.
- [9] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. A. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, *The HTK Book (for HTK Version 3.4)*, Dec. 2006.

Domain Analysis and Feature Modeling of GUIs

¹Michaela BAČÍKOVÁ (2nd year)
Supervisor: ²Jaroslav PORUBÁN

^{1,2}Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹michaela.bacikova@tuke.sk, ²jaroslav.poruban@tuke.sk

Abstract— Feature modeling is a technique used for software product lines to create models of common and different features in the software products. There are many methods for feature modeling, each of them however focus mainly on the modeling process and on the use of the resulting models in generative processes and only a few of them deal with the process of domain modeling, which is the first step for creating a feature model. In the field of domain modeling there are automatized methods for creating domain models, they however create these models from source codes, databases and documents written in natural language. We think that the best source of formalized domain information is graphical users interface (GUI) because it is primarily intended for a domain user. Although it is in a formalized form, GUI is not immediately usable as a domain model because of many implementation details involved. Therefore we try to propose a method for automatic domain analysis of GUIs driven by stereotypes of creating GUIs, AGDAM. For notation of domain models we want use FeatureIDE notation. In this article we will describe basic processes for creating such feature models from GUI component examples.

Keywords— feature modeling, domain analysis, graphical user interface, components

I. INTRODUCTION

Domain analysis (DA) is a process that is currently most often used in the software systems design and analysis. The Fig. 1 describes the most common process of domain analysis. It is necessary to collect information from various sources: a) **domain users and experts**; b) **existing documents**; and c) **existing software systems**. Then the different kinds of information are categorized, analyzed and some forms of a

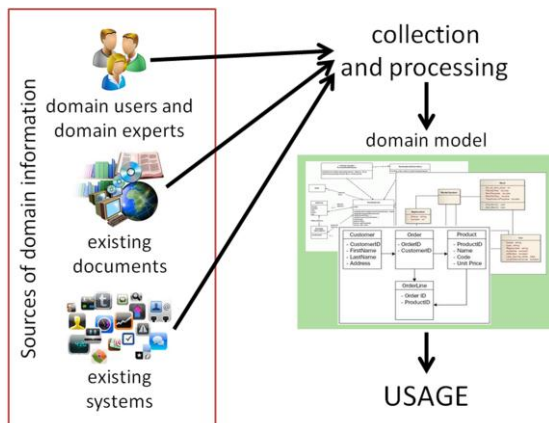


Fig. 1. An illustration of the domain analysis process.

domain models are created.

The domain models have many forms. Commonly used is

the feature model in different notations (or example original FODA notation described in [1] or Czarnecki-Eisenecker notation described in [2]). Feature modeling methods analyze product lines and resulting domain model contains varying and consistent features in a domain and also defines the vocabulary used in the domain, which defines concepts, ideas and phenomena within the system. This model (often with a connection with generators and libraries of reusable components or frameworks) is used for creating a new software system or editing an existing system. The domain analysis is often performed by a **domain analyst**.

II. PROBLEM STATEMENT

Modeling and creating domain models from gained information is now widely supported by numerous methodologies and tools. Gaps remain especially in the areas of **data collection**. Problems of existing approaches in this area can be divided into three groups based on the information source. Each group will be analyzed in the next chapters.

A. Problems in the area of data collection from domain users and domain experts

Data collection from domain users and experts is often carried out through reviews, questionnaires and forms. These methods are often *time consuming* and require both *willingness of users and experts* and a certain level of *skill of the domain analyst*. On one hand, a user or a domain expert do not always have the time or mood to fill out some questionnaires or talk with analyst. The domain analyst on the other hand must be able to ask the right questions to gain the information he needs.

B. Problems in the area of data collection from non-formalized existing documents

Various automatic techniques are used. The most common are NLP (Natural Language Processing) methods and techniques of AI (Artificial Intelligence). Sometimes also an existing design documentation of software systems is analyzed. The biggest problem of these approaches is the ambiguity of the natural language and therefore there is always a need for additional control and adjustments by a human expert. Finally, it is necessary to gain existing documents to be analyzed and they may not even exist for the particular domain.

C. Problems in the area of data collection existing software systems

The approaches for automatic collection of domain

information from existing software systems mainly use static analysis of source codes and databases. Databases or source codes are however *not primarily intended for the user*. Hence the author of database or code is **not forced to use domain vocabulary** during development. So the domain terms may not even be included in the database or source code analyzed, they can be written in another language, abbreviations can be used, or there may be some other language barrier.

The analysis of databases is very easy: the names of tables and relations between tables can be exactly determined and represent domain terms and relations. There is, however, *no description of domain processes*.

On the other hand, source codes contain descriptions of domain processes in a form of methods or functions. There is however a *high level of implementation details* getting in the way, preventing the domain analysis to be carried out. *Generalization*, which is currently widely used in the implementation of reusable systems, represents another barrier. The aim of generalization is to use generic terms that can be used to describe objects and thus to ensure reusability of the system also in other domains. This may hinder the domain analysis: a domain-specific model should contain specific terms, not general.

III. PROPOSED SOLUTION

There are many methods and tools supporting communication with users and domain experts. Also in the area of domain analysis of non-formalized documents there is a significant number of research papers proposing different methodologies and techniques of NLP and AI (see chapter V.). Since these two areas are more or less exhausted, we decided to focus on the last area, the analysis of existing software systems, as this area is only little explored in relation to domain analysis.

Based on the problems and facts identified in the previous chapter we think that more appropriate target for the DA is a **user interface**. A user who comes from the target domain has a *direct access to the UI*. For the user to be able to use the system effectively it must be built with respect to *understandability*, i.e. it must contain *terms from the domain*. It also should describe *domain processes* in a form of event sequences executed on the UI.

Currently applications are often created as component-based because of reusability support. In component-based applications it is possible to separate source code components from user interface components to a certain level. Although there are still very unclear boundaries between application presentation layer and source code layer. Using reflection and aspect oriented programming allows us to separate the implementation details from domain information that we need.

In this paper we will describe the basic graphical components and a method for formalizing them (or their groups) into a domain model illustrated on examples. This will represent a basis for our future research in the area of automated GUI domain analysis.

A. Previous research

This paper is based on our previous work in this field, see [3], [4]. In our recent work [5] we analyzed basic stereotypes of creating user interfaces and based on them we defined basic

rules for GUI formalization that will be used in implementation of our Automatic GUI Domain Analysis Method (AGDAM). We created an automated tool for extracting information from UI and automated recorder of user actions performed on UI. This tool we want to use in our further research in the area of DA.

IV. MOTIVATION

Besides the tasks of creating a new software system, or editing an existing software system, the resulting domain model can be used also in other areas, for which we primarily aim our future work. The domain model can be used for example in model driven engineering (MDD) for *generating various software artifacts or whole software systems*. For example Ristic *et al.* [6] use form specifications to generate forms, however when using their tool the user alone must enter the form specifications into the computer. An automatic domain analysis of existing systems with forms (for example web sites) for such form specifications could help the user with formalizing the requirements and the user would only edit the results of the DA according to his needs.

If there already is an existing software system, based on the domain model of such system and with the help of dynamic analysis of user handling this system and generators, *documentation* (for example a user guide) and different types of *models* (for future development) can be generated.

Last but not least area of use is an automatic evaluation of *domain usability*. Domain usability is a usability related to domain terms contained in the interface. For the user interface to be usable in a domain the domain user must understand it, we say that it must be *understandable*. With our method, it is possible to extract all terms located in the interface and analyze them for membership in the domain. If there are terms, that do not belong to the domain (for example in an internet banking system, there should not be terms like “student”, “lecturer” and so on), then they represent a domain usability issue.

V. STATE OF THE ART

The domain analysis was first defined by Neighbors [7] in 1980. By introducing the Draco approach, a code generator system that works by integrating reusable components Neighbors stresses, that domain analysis is the key factor for supporting reusability of analysis and design.

Widely used approach for domain analysis is the FODA (Feature Oriented Domain Analysis) approach [1], that aims for analysis of software product lines by comparing the different and similar features or functionalities of systems. The method explains what the outputs of domain analysis are but remains vague about the process to obtain them. Very similar to the FODA approach and practically based on it, is the DREAM (Domain Requirements Asset Manager) approach by Mikeyong *et al.* [8]. They also perform commonality and variability analysis of product lines, but with the difference of modeling domain requirements, not features or functionalities of systems. Many approaches and tools support the FODA method, e.g. Ami Eddi [9], CaptainFeature [10], RequiLine [11], ASADAL [12], pure::variants [13].

Not only the process of DA is supported by some

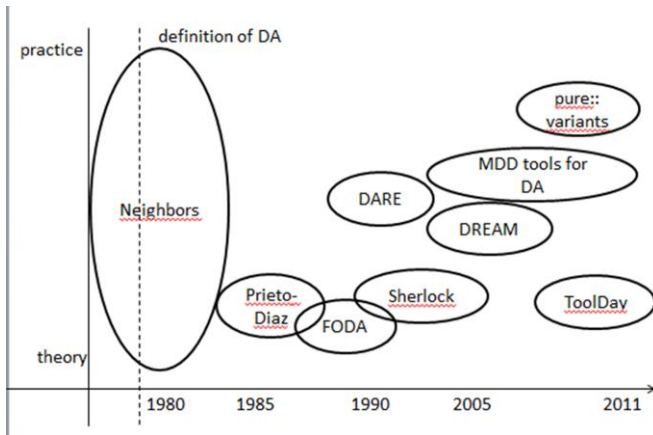


Fig. 2. A time chart of existing approaches and tools for domain analysis (inspired by the time chart by Frakes *et al.* [20])

approaches, but also reusability feature by providing a library of reusable components, frameworks or libraries. Such approaches are for example the early Prieto-Diaz approach [14] that uses a set of libraries; or the later Sherlock environment [15] that uses a library of frameworks.

The latest efforts are in the area of MDD (Model Driven Development). The aim of MDD is to shield the complexity of the implementation phase by domain modeling and generative processes. The MDD principle support for example the Czarnecki project Feature Plug-in [16] or his newest effort Clafer [17] and an eclipse plug-in FeatureIDE [18] by Thüm and Kästner.

ToolDay (A Tool for Domain Analysis) [19] is a tool with support for every phase of DA with possibilities for validation, generating models and exporting to different formats.

All tools and methodologies listed here support the DA process by analysis of data, supporting summarizing, clustering of data, or modeling features. But the input for domain analysis (i.e. the information about the domain) always come from the users, or it is not specified where they come from. Only the DARE (Domain analysis and reuse environment) tool from Prieto-Díaz [20] primarily aims for *automatic* collection and structuring of information and creating a reusable library. The data are collected not only from human resources, but also automatically from *existing source codes and text documents*. But as mentioned above, the source codes do not have to contain the domain terms and domain processes. The DARE tool *does not analyze the UIs* specifically.

A time chart of the approaches can be seen on Fig. 2.

VI. DOMAIN MODELING OF BASIC GRAPHICAL COMPONENTS

The fundamental building block of each GUI is usually a window or screen, to which all components are put into. It represents the domain of the application and the title of the window represents the name of the domain. The basic components which we will deal with in this paper are: windows, containers, tabpanes, buttons (classic buttons, radio buttons, checkboxes).

We use FeatureIDE notation for our examples of domain models because of their easily readable form and very clear representation of hierarchical relations. Hierarchical relations in the model examples can be derived from component structure. These models can be used in feature modeling. The basic domain dictionary can be created automatically and the



Fig. 3. GUI of iAlertU

user can edit the feature model himself after its automatic generation.

On Fig. 3 there is a simple GUI of an open source Java application iAlertU, which we choose for this experiment. The

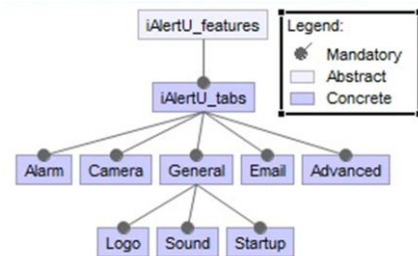


Fig. 4. Basic domain model derived from iAlertU GUI

GUI is simple, it uses only one window and tabpanes are used for representing its basic structure.

As we determined in our previous works, the hierarchy of the basic terms can be derived from the container structure. In this case it is the structure of tabs and containers in them. On Fig. 4 there is a basic domain model derived from iAlertU GUI based on the structure of tabs. The name of the main feature (*iAlertU_features*) is derived from the title of the window. The names of features are derived from the tab titles.

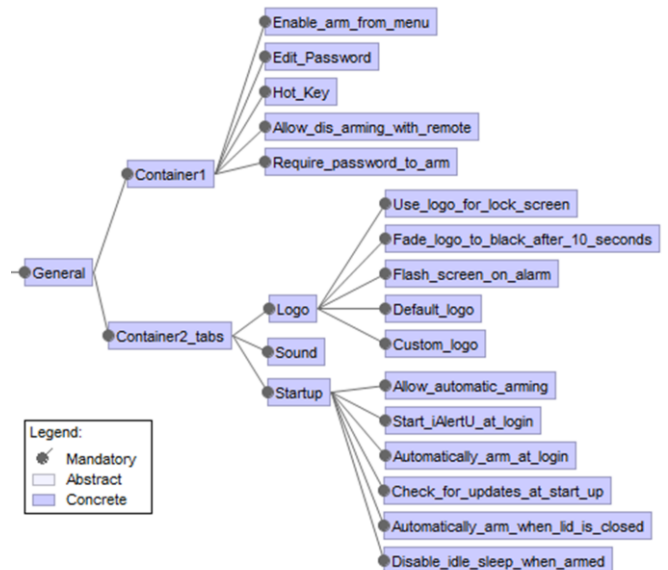


Fig. 5. Extended feature model derived from iAlertU GUI

All these properties can be programmatically determined by our approach.

This however can be taken even further. To address all functionalities of the window, we can add every feature of the GUI to the model. In our approach every component displayed in the UI can be identified in the component tree

automatically.

The domain descriptors (for example the domain descriptor of the General tab is the “General” string, i.e. the string which is visible in the UI), used in the model are derived from the component attributes, such as *title*, *labelFor* attribute of a Label component, *toolTipText*, etc.

On Fig. 5 there is an example of extended feature model derived from the iAlertU UI. To simplify it, we displayed only a part of the whole model, the General tab.

We have already implemented an automatic method for analyzing graphical user interfaces of Java applications, which is able to derive the component tree and also to identify components according to their domain descriptions. Our next step is to design and implement an algorithm for automatic derivation of feature models from such component tree.

VII. CONCLUSION

In this paper we described the theory behind domain analysis and we identified the main gaps in the DA of software systems. We determined that the best possible source of information is the presentation layer of software systems, because it is already formalized and it is close to the user (and therefore also close to the target domain). We provided analysis of the actual state of the art and an insight of our previous and recent research. We outlined our aim of designing and implementing the AGDAM method that is a center of our current and future research and we proposed ways of extracting information from various graphical components that will serve for implementation of the AGDAM method, which we illustrated on examples.

ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/0305/11 – Co-evolution of the artifacts written in domain-specific languages driven by language evolution.

REFERENCES

- [1] K C Kang, S G Cohen, J A Hess, W E Novak, and A S Peterson, "Feature-Oriented Domain Analysis (FODA) Feasibility Study," Tech. Rep. CMU-SEI-90-TR-21, 1990.
- [2] Krzysztof and Eisenecker, Ulrich W. Czarnecki, Generative programming: methods, tools, and applications. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000.
- [3] Michaela Bačíková, Jaroslav Porubán, and Peter Václavík, "First Step for GUI Domain Analysis: Formalization," Journal of Computer Science and Control Systems, vol. 4, no. 1, pp. 65–69, 2011.
- [4] Michaela Bačíková, Jaroslav Porubán, and Dominik Lakatoš, "Introduction to Domain Analysis of Web User Interfaces," in Proceedings of the Eleventh International Conference on Informatics (Informatics'11), Rožňava, 2011, pp. 103-108.
- [5] Michaela Bačíková and Jaroslav Porubán, "Analysing Stereotypes of Creating Graphical User Interfaces," Central European Journal of Computer Science, submitted for publication 2012.
- [6] S., Aleksic, S., Lukovic, I., Banovic, J. Ristic, "Form-Driven Application Generation: A Case Study," in Proceedings of the Eleventh International Conference on Informatics, INFORMATICS'2011, Rožňava, Slovakia, 2011, pp. 115-120.
- [7] James M Neighbors, Software Construction Using Components, 1980, Dissertation, Department of Information and Computer Science, University of California, Irvine.
- [8] Mikyeong Moon, Keunhyuk Yeom, and Heung Seok Chae, "An Approach to Developing Domain Requirements as a Core Asset Based on Commonality and Variability Analysis in a Product Line," IEEE Trans. Softw. Eng., vol. 31, no. 7, pp. 551–569, July 2005.
- [9] Bendash Thomas, Unger Peter, Eisenecker Ulrich Czarnecki Krzysztof, "Generative programming for Embedded Software: An Industrial Experience," in Generative programming and component engineering, ACM SIGPLAN/SIGSOFT Conference (GCPE 2002) Proceedings, LCNS 2487, vol. 1, Pittsburgh PA, USA, 2002, p. 164.
- [10] CaptainFeature website on SourceForge.net, last actualisation 2005. [Online]. <https://sourceforge.net/projects/captainfeature>
- [11] RequiLine website, last actualization 2005. [Online]. <https://www-lufgi3-informatik.rwth-aachen.de/TOOLS/requiline/index.php>
- [12] Review of ASADAL CASE framework, Postech Software Engineering Laboratory, last actualization 2011. [Online]. http://selab.postech.ac.kr/ASADAL-Simple_Overview.pdf
- [13] Otto-von-Guericke-Universität Magdeburg and Fraunhofer Instituts Rechnerarchitektur und Softwaretechnik. (2011) pure:variants. http://www.pure-systems.com/pure_variants.49.0.html.
- [14] R. Prieto-Díaz, "Reuse Library Process Model," Electronic Systems Division, Air Force Systems Command, USAF, Hanscom AFB, AM, STARS Reuse Library Program Contract F19628-88-D-0032, Task IS40, 1991.
- [15] Andrea Valerio, Giancarlo Succi, and Massimo Fenaroli, "Domain analysis and framework-based software development," SIGAPP Appl. Comput. Rev., vol. 5, no. 2, pp. 4--15, September 1997.
- [16] Michal Antkiewicz and Krzysztof Czarnecki, "FeaturePlugin: feature modeling plug-in for Eclipse," in Proceedings of the 2004 OOPSLA workshop on eclipse technology eXchange, eclipse '04, Vancouver, British Columbia, Canada, 2004, pp. 67--72.
- [17] Kacper Bąk, Krzysztof Czarnecki, and Andrzej Wasowski, "Feature and meta-models in Clafer: mixed, specialized, and coupled," in Proceedings of the Third international conference on Software language engineering (SLE'10), Eindhoven, The Netherlands, 2011, pp. 102--122.
- [18] Thomas Thüm, Christian Kästner, Sebastian Erdweg, and Norbert Siegmund, "Abstract Features in Feature Modeling," in 15th International Software Product Line Conference (SPLC), 2011, pp. 191--200.
- [19] Liana Lisboa, Vinicius Garcia, Eduardo de Almeida, and Silvio Meira, "ToolDay: a tool for domain analysis," International Journal on Software Tools for Technology Transfer (STTT), vol. 13, no. 4, pp. 337-353, 2011.
- [20] William Frakes, Ruben Prieto-Diaz, and Christopher Fox, "DARE: Domain analysis and reuse environment," Ann. Softw. Eng., vol. 5, pp. 125--141, January 1998.

Anchor-free positioning using ToA node distance estimation and LoS/NLoS detection

Vladimír CIPOV (3rd year), Marek COPÁK
Supervisor: Ľubomír DOBOŠ

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

vladimir.cipov@tuke.sk, marek.copak@student.tuke.sk, lubomir.dobos@tuke.sk

Abstract—The problem of the node localization in non-infrastructure Mobile Ad hoc NETWORKS (MANETs) is presented. We present an Anchor-free positioning algorithm. It consists of two stages. At first the Multidimensional Scaling (MDS) technique is used to form node local coordinate systems which are unified in the second stage to one network coordinate system. Time of Arrival (ToA) ranging technique is utilized in order to determine the node distance, estimated from the channel impulse response (CIR). In order to increase the accuracy of the ToA parameter estimation, the analysis of the wireless communication channel behaviour is applied using the kurtosis parameter calculation. It has been observed how the localization accuracy will be influenced in the case that it is possible to estimate the channel behaviour.

Keywords—positioning, LoS/NLoS detection, anchor (GPS)-free, Time of Arrival, kurtosis index.

I. INTRODUCTION

In the field of mobile communications, the wireless nodes location problem has gained increasing interest. Several positioning algorithms have been developed so far, but none of them is applicable for localization in a random environment. Each algorithm has just very narrow area of application. The first common classification of the localization algorithms according to the reference nodes availability is to Anchor-based and Anchor free. The second common is according to the requirement of the signal parameter measurement to Range-based and Range-free [8]. In this paper we combine the Anchor-free and Range-based localization principle that is based on the Improved MDS-MAP algorithm [1] and ToA ranging technique [6]. One of the main drawbacks of the Anchor-free localization is that the location of an unknown node is set relative to the unstable coordinate system. The most commonly used ranging techniques are Received Signal Strength (RSS), Time of Arrival (ToA), Time Difference of Arrival (TDoA) and Angle of Arrival (AoA). The ToA is one of the most accurate ranging techniques, however, it suffers problems of smaller accuracy especially under multipath environment, and its implementation relates with higher costs and computational load.

The remainder of this paper is structured as follows. In section II the anchor-free positioning algorithm is described. Section III contains the wireless channel classification. The use of the kurtosis parameter, as the parameter for channel

identification, is explained in the section IV. The simulations and results are summarized in section V and concluded in section VI.

II. BRIEF OVERVIEW OF THE IMPROVED MDS-MAP SELF POSITIONING ALGORITHM

Improved MDS-MAP algorithm [1] runs for each node separately and uses only one-hop communication, which prevents the network congestion. It consists of two main steps:

1. **Local coordinate system (LCS) formation:**
 - a. node distance estimation based on the localization parameter measurement,
 - b. distance matrix formation.
2. **Global coordinate system (GCS) formation.**

A. Local coordinate system formation

The aim of this process is LCS formation of each node in the network, which consists of its relative position and relative position of its one-hop adjacent nodes. Each node forms its own coordinate system, in which it is located at the centre.

Node distance estimation based on localization parameter measurement. Some distance estimation techniques, such as RSS or ToA, can be used in order to determine the node distance. Now, each node knows its distance to the one-hop adjacent nodes and fills the first row and column of the distance matrix (Fig. 1a, node number one forms its distance matrix. Its LCS consists of the nodes number 1, 3, 7, 8, 9).

Distance matrix formation. In this step the missing node distances are filled up in the distance matrix. This process is based on the node communication procedure, Fig. 2, of such nodes, which belong into the LCS of the concrete node (in this case it is the node number one).

ID	1	3	7	8	9
1	0	5	10	2	9
3	5	0	-	-	-
7	10	-	0	-	-
8	2	-	-	0	-
9	9	-	-	-	0

a

ID	1	3	7	8	9
1	0	5	10	2	9
3	5	0	7	6	15
7	10	7	0	3	9
8	2	6	3	0	4
9	9	15	9	4	0

b

Fig. 1. Distance matrix formed in the 1st step (a) and in the 2nd step (b)

The node “ID1” sends a message containing its identification to the one-hop adjacent nodes in the network.

They add the information to the message about the distance to the node "ID1" and send this message to their one-hop adjacent nodes. This procedure is repeated also by the third degree nodes. They add the information to the message about their distance to the node, from which they receive the message. These nodes send the message again to their adjacent nodes and they check:

- If "ID1" is identical with node ID, it stores the information about the distance between the nodes "ID2" and "ID3" into its distance matrix.
- For other nodes, if "ID1" is not identical with node ID, ignore this message and do not send it again.

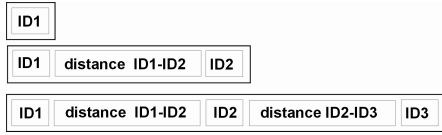


Fig. 2. Example of the message sent by node ID1 to the node ID3

The distance matrix is symmetrical and contains zeros in the main diagonal. The formed distance matrix is used as the input into the Multidimensional Scaling method in order to form a local coordinate system of each node.

B. Global coordinate system formation

In the second step, the partial LCSs must be unified, because they are shifted, rotated and mirrored according to each other. Using such operations, the partial LCSs can be joined into the one global coordinate system. The procedure of the GCS formation is explained on two LCSs of the nodes "A" and "B". For the sake of clarity, the nodes of the second LCS are marked by a comma, Fig. 3. The LCS of the node B (LCS(B)) is joined with the LCS of the node A (LCS(A)). Presence of the node "B" located in the LCS(A) (and vice versa, the node "A" located in the LCS(B)) is crucial. Similarly, it is advantageous, if the third node is located in LCS(A) and LCS(B) simultaneously.

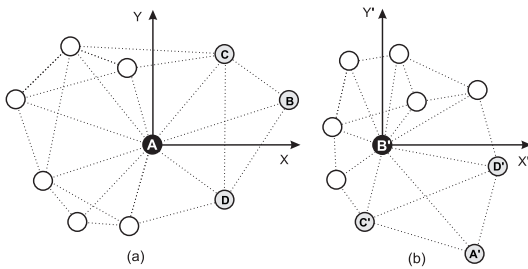


Fig. 3. LCSs of the nodes "A" and "B" used for problem explanation

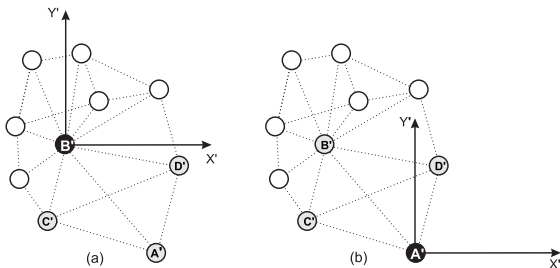


Fig. 4. LCS(B) shifted to the node "A", (a) before shifting, (b) after shifting

At first the centre of the LCS (B) must be shifted to the node "A",

Fig. 4. The procedure of shifting is carried out by subtraction of the coordinates of the node "A" from the coordinates of each node.

Next, both LCSs must be oriented in the same direction, in other words, LCS(A) must be rotated by an angle α and LCS(B) must be rotated by an angle β . They can be expressed as, Fig. 5:

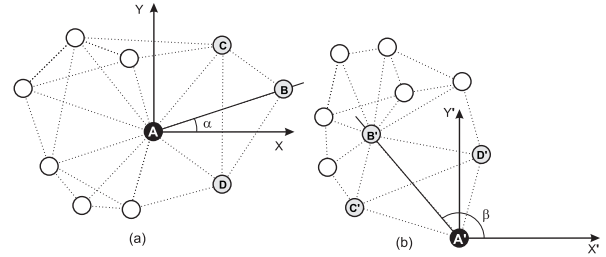


Fig. 5. Calculation of angles for rotation process

$$\alpha = \arccos \frac{x_B}{D_{AB}}; \quad \beta = \arccos \frac{x'_B}{D'_{AB}} \quad (1)$$

for positive y_B, y'_B and

$$\alpha = -\arccos \frac{x_B}{D_{AB}}; \quad \beta = -\arccos \frac{x'_B}{D'_{AB}} \quad (2)$$

for negative y_B, y'_B .

$[x_B, y_B]$ represent the coordinates of the node "B" and D_{AB} represents the distance between the nodes "A" and "B". Similarly for $[x'_B, y'_B]$ and D'_{AB} .

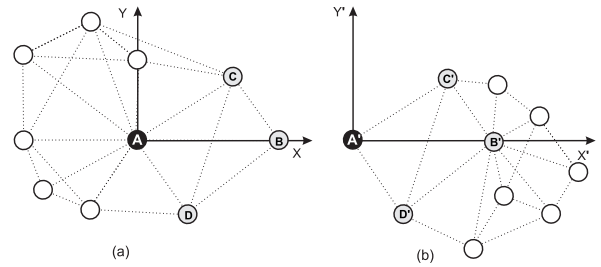


Fig. 6. LCSs oriented in the same direction

Then both LCSs will rotate in the direction negative to the size of α or β . For this process it is necessary to calculate the angle δ , which represents the angle BAX_i (X_i represent i^{th} node; $X_i \neq A, B$).

For positive and negative y_i respectively:

$$\delta_i = \arccos \frac{x_i}{D_{Ai}}; \quad \delta_i = -\arccos \frac{x_i}{D_{Ai}} \quad (3)$$

The point coordinates will be recalculated as:

For positive and negative result of $\delta - \alpha$ respectively:

$$\begin{aligned} x_i &= D_{Ai} \cos(\delta_i - \alpha); & y_i &= D_{Ai} \sin(\delta_i - \alpha) \\ x_i &= -D_{Ai} \cos(\delta_i - \alpha); & y_i &= D_{Ai} \sin(\delta_i - \alpha) \end{aligned} \quad (4)$$

The same procedure applies also for the LCS (B) rotation.

After rotation, a point “C”, which is common for both LCSs, must be found. According to its location, it will be decided about the need of mirroring. If “C” has not the same sign of the y-axis, mirroring is necessary, Fig. 7.

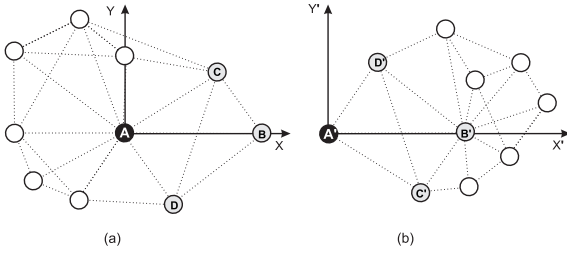


Fig. 7. Test for the need of mirroring (now mirroring is needed)

Mirroring is carried out by simple change of the sign of the y-axis for each node in LCS. Now both LCSs can be joined together, Fig. 8.

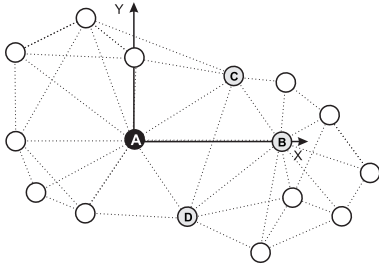


Fig. 8. LCS(B) joined to the LCS(A)

III. WIRELESS COMMUNICATION CHANNEL CLASSIFICATION

The channel impulse response between a transmitter and a receiver distant at distance d can be expressed as [2]:

$$h_d(t) = \sum_{i=1}^{L_p} \beta_i^d \delta[t - \tau_i^d] \quad (5)$$

where L_p is the number of multipath components, β_i^d is the location specific random complex amplitude and τ_i^d represents the random propagation delay of the i -th path. Impulse transmitted from the transmitter site arrives at the receiver site as the sum of multiple delayed impulses δ with different magnitude and phases due to multipath behaviour of the wireless communication channel. Wireless communication channel profile can be classified as [2]:

- **LoS (DDP) – Line-of-Sight with Dominant-Direct-Path.** In this case the direct path is the strongest and first detected path in the channel profile and it can be detected by the receiver. There are no obstacles between the transmitter and the receiver.

For range measurements in NLOS channel conditions face a more difficult challenge, since the direct path is either detected but attenuated or completely blocked [3].

- **NLoS (NDDP) – Non Line-of-Sight with Non Dominant-Direct-Path.** Line-of-Sight communication between a transmitter and a receiver does not exist. In NDDP channel profile the direct path is not the dominant strongest path of the channel impulse response, however, it is detectable because it is

received above the detection threshold of the receiver. Direct path can be detected by a more complex receiver. It may happen that the receiver chooses the direct path incorrectly according to path with strongest energy.

- **NLoS (UDP) – Non Line-of-Sight with Undetected-Direct-Path.** In UDP channels, direct path arrives below the detection threshold of the receiver while other paths are still detectable. The receiver assumes the first detected peak as a direct path which causes considerable distance estimation errors.

IV. KURTOSIS INDEX – PARAMETER FOR CHANNEL IDENTIFICATION

Knowing the channel profile is very important in order to maximize performance of the localization algorithms working in wireless communication environment especially in combination with ToA distance estimation technique. In the literature there are few papers exploring parameters that allow LoS or NLoS identification [4][5]. One of them is Kurtosis index [6]. The kurtosis is a statistical parameter that indicates the variance variation of the signal amplitudes. Thus, it indicates with high numerical values the signals with distinguishable amplitude peaks. Kurtosis is mathematically defined as follow:

$$kurtosis(x) = \frac{1}{\sigma^4} \frac{\sum_i (x_i - \bar{x})^4}{N} \quad (6)$$

where σ is the standard deviation of the variable x and \bar{x} is the mean value of x . N is the number of samples of x .

Kurtosis index is characterized by the high value of the signal received via LoS communication (few peaks and very distinguishable from natural noise), and by low value of signal received via NLoS communication (the signal is more noise-like). In general, kurtosis has a smaller value if SNR of the received signal becomes smaller [6].

Consequently, kurtosis parameter calculated from the channel impulse response is compared with the threshold value defined for individual environment. It is clear, that it takes different values for different environments. In general, if the calculated values of the kurtosis parameter take the values above the defined threshold, it indicates the LoS communication and vice versa, if the calculated values of the kurtosis parameter take the values less than the defined threshold, it indicates the NLoS communication.

Another advantage of kurtosis is that it is calculated directly from the received signal, no other estimation algorithm is needed.[6].

V. SIMULATIONS AND RESULTS

The described anchor-free positioning algorithm is evaluated in the following simulations. It is the Range-based positioning algorithm therefore it is combined with ToA node distance estimation technique. In order to simulate the real wireless communication, the existing model of the 802.15.4a

technology has been used, which is described in [7]. It provides the wireless channel impulse response generation for different types of environment. For the simulations, the office environment has been chosen, in which the LoS and NLoS communication conditions have been randomly generated. The communication channel type and the real node distance have been the input parameters to the CIR generation function.

Three ToA estimation algorithms have been compared:

- **strongest peak algorithm** - the ToA parameter is determined in accord with the time of arrival of the received path represented by the strongest peak.
- **combination of the strongest peak and first detected peak algorithms** in accord with the calculated kurtosis parameter - KP (LoS/NLoS detection is used) - if the communication is classified as LoS, the ToA parameter is determined as the time of arrival of the strongest path. If the communication is classified as NLoS, the ToA parameter is determined by the time of arrival of the first detected path.
- **first detected peak algorithm** – it does not distinguish between the LoS and NLoS communication. The ToA parameter is always determined by the time of arrival of the first detected peak.

The detection threshold of the kurtosis parameter for LoS/NLoS identification was obtained from the experimental results as an average value, where the kurtosis parameter was calculated from the thousand times generated channel impulse response for LoS and NLoS office environment separately. It was set on **112** ($KP \geq 112$ indicates LoS and on the other hand $KP < 112$ indicates NLoS). The next table shows the successfulness of the channel identification in the simulations:

TABLE I
PRECISION OF THE LOS/NLOS CHANNEL IDENTIFICATION

Type of environment	Successfulness of the LoS/NLoS identification [%]
Office	82,50

If the ToA parameter is known, the node distance D can be simply expressed as:

$$D = c \cdot ToA \quad (7)$$

where c represents the speed of light ($3 \cdot 10^8$ m/s) and ToA represents the estimated ToA parameter.

The simulations were carried out in MATLAB programming environment and for statistical relevance they were repeated one hundred times. We used the network which consisted of the 50 mobile nodes deployed on the area of 50x50m with all receivers' radio range of 15m. We assumed the precise transmitter-receiver time synchronization. Some parameters of the localization process have been observed such as the localization error, the number of successfully localized nodes and the node distance estimation error, respectively. The localization accuracy and distance estimation error were expressed in percentage according to the network size (50m=100%). The reader can find the results of the simulations in the TABLE II.

TABLE II
POSITIONING ALGORITHM EVALUATION FOR EACH INDIVIDUAL CASE OF USED TOA NODE DISTANCE ESTIMATION ALGORITHMS

ToA estimation algorithm	Total localization error [%]	Average number of localized nodes	Average distance estimation error [%]
Strongest peak	18,97	48,48	2,47
LoS/NLoS detection	14,21	49,00	0,75
First detected peak	8,49	48,42	0,22

VI. CONCLUSION

In this paper the anchor-free localization was explained using the MATLAB simulation of the Improved MDS-MAP positioning algorithm in combination with Time of Arrival ranging technique. The problem of precise time delay estimation from the received signal has been introduced especially in harsh environment. It has been shown that knowledge of the wireless channel behaviour can significantly affect the node distance estimation and node location accuracy.

ACKNOWLEDGMENT

This work has been performed partially in the framework of the EU ICT Project INDECT (FP7-218086) (40%) and by the Ministry of Education of Slovak Republic under research VEGA 1/0386/12 (30%). This work is also the result of the project implementation Development of the Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (30%).

REFERENCES

- [1] Y. Shang, W. Ruml, "Improved MDS-Based Localization," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 15 (11), November 2004, pp. 961-974.
- [2] A. Hatami, "Application of Channel Modeling for Indoor Localization Using TOA and RSS" Dissertation thesis. Worcester Polytechnic Institute, May 2006.
- [3] N. Alsindi, D. Chunjie, J. Zhang, "NLOS Channel Identification and Mitigation in Ultra Wideband ToA-based Wireless Sensor Networks" in *Proceedings of 6th Workshop on Positioning, Navigation and Communication (WPNC)*, March 2009.
- [4] A. Ciurana, F. Barceló, S. Cugno, "Multipath profile discrimination in ToA-based WLAN ranging with link layer frames" *ACM Int. Workshop on Wireless Network Testbeds, Experimental evaluation and Characterization (WiNTECH 2006)*, pp. 73-79, Los Angeles (USA), Oct. 2006.
- [5] M. Heidari, A.F. Ozan, K. Pahlavan, "Identification of the absence of direct path in indoor localization systems" In *International Journal of Wireless Information Networks*, Worcester: Centre for wireless information network studies, Worcester Polytechnic Institute, 2008.
- [6] L. Mucchi, P. Marcocci, "A new UWB indoor channel identification method" in *the Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, Orlando, Florida, USA, July 31 - August 3, 2007.
- [7] A.F. Molisch, et al., "IEEE 802.15.4a Channel Model – Final Report" Tech. Rep., Document IEEE 802.1504-0062-02-004a, 2005.
- [8] 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.

Current Trends in Human–Robot Interaction: Towards Collaborative & Friendly Machines

¹Mária VIRČÍKOVÁ (2nd year), ²Peter SMOLÁR (3rd year)
Supervisor: ³Peter SINČÁK

^{1,2,3}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹maria.vircikova@tuke.sk, ²peter.smolar@tuke.sk, ³peter.sincak@tuke.sk

Abstract— The focus of robotic research continues to shift from industrial environments, where robots perform repetitive tasks, to service robots placed in a wide variety of environments, often in human-habited ones. Thus, the importance of the field of Human-Robot Interaction (HRI) has been denoting an increasing growth. This paper identifies current trends in the research of Human-Computer Interaction, focusing on new kinds of interfaces, as it has many overlapping challenges and applications with HRI. We also discuss the question whether robots are a distinctive case of study than other technologies studied in HCI. We believe that the quality of HRI will determine the effectiveness of the collaboration and in general, the acceptance of robots in the society. Our main research is focused on developing computational models for humanoid robots which try to predict what a person is thinking or wanting. We illustrate this idea on two case studies: the first one is an architecture based on the robot IROS, endowed with a system for prediction of human's goals in a collaboration task and the second is an architecture based on the robot NAO, possessing an emotion technology, which is used for guessing user's future behaviour based on his/her emotional expressions. We consider this approach useful to make the HRI more natural and efficient.

Keywords—human-computer interaction, human-robot interaction, social robotics, robot IROS, robot NAO.

I. INTRODUCTION

*“Above all, the human animal is social.
For an artificially intelligent system, how could it be otherwise?”*

As robots increasingly make their way into functional roles in human environments (e.g. homes, schools, and hospitals), they need to react appropriately to human expectations and behaviour. Moreover, a person working with a robot should not be required to learn a new form of interaction. Thus, we need to develop computational models of social intelligence for these robots that will allow them to have interactions that are natural and intuitive for a human partner.

According to the Human-Robot Interaction (HRI) Research Portal for the HRI Community [2] “HRI is a field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans”, including topics of gesture and natural language communication, perceiving emotions and expressing their own artificial emotions, establishing social relationships, exhibition of different personalities and characters, recognition of interaction partner and others.

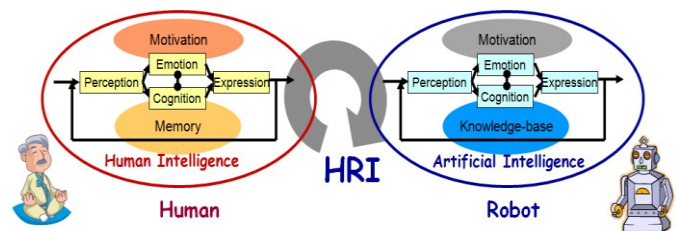


Figure1. HRI is linkage between these two agents with different levels of perceptual/cognitive/emotional/expressional abilities according to HRI Group at KAIST[3].

Interaction requires communication between robots and humans, which may be remote (teleoperation/telemanipulation), proximate (may include a physical interaction, e. g. robot assistant) and social (including social, emotive, and cognitive aspects).

The field of cooperative work sees robots and humans acting like team members rather than autonomous, independent devices. If an intelligent behaviour is required, devices should be endowed with metaknowledge and meta-communicative skills, e.g. consciousness and self-awareness. Current trend is to involve emotion technology in such systems—it can help to determine goals and communicate internal states of humans and robots. A number of cognitive scientists believe that intelligent behaviour cannot take place without emotion [11].

This paper has the following contributions:

In Section II we try to draw attention to the new generation of Human-Computer interfaces that can be useful for developing more efficient HRI presenting a survey of them.

Section III develops an idea of human-oriented robotics, where we consider the emotional technology as a new trend in such systems.

Section IV presents experiments with cognitive robots in the human-robot collaboration tasks.

Researchers want to construct robots to behave more like people, so that people do not have to behave like robots when they interact with them. The global idea is that communication and interaction should be easy and enjoyable, both for unfamiliar users and trained professionals.



Figure 2. From left: I-Garment is developing full-bodied smart garments – to be worn by fire-fighters and the like – that monitor and transmit the location and vital signals of its wearer (such as body temperature and heartbeat).

Microsoft's 'Surface' is an interactive tabletop allowing two-handed interaction with digital objects such as photos, music files, games and maps. These kinds of interactive surfaces encourage collaborative, creative engagement. 0

g.MOBilab is a super-low-weight biosignal recorder which measures EEG, data processing, analysis, and pattern recognition to create a fully functional Brain Computer Interface

II. NEW TRENDS IN HUMAN-COMPUTER INTERACTION

As the field of HRI is growing, it benefits (in methodologies, design principles, and computing metaphors) from researchers in HCI (and vice versa) and it has been nurtured by HCI organizations.¹ This field is still a young one - sometimes considered a specialized offshoot of the wider area of human-computer interaction (HCI), other times a new field - we point out some of many contributions it has from HCI.

Microsoft Research in 0summarises major transformations that affect how people will interact with computing technology in future. They argue that computing no longer has a single interface (as the conventional computer with a keyboard and a mouse), but many different ones - some are created by computers encroaching people's personal space, even being embedded within them, others are produced by computers moving away and disappearing into the richness and complexity of the world around them.

A. Transformation of HCI interfaces

As a new trend can be considered **Intelligent HCI designs**-interfaces that incorporate at least some kind of intelligence in perception from and/or response to users, e.g. speech enabled interfaces that use natural language to interact with user or devices that visually track user's movements or gaze and respond accordingly. (Adaptive HCI designs, on the other hand, may not use intelligence in the creation of interface but use it in the way they continue to interact with users.)

From GUIs to multi-touch, speech to gesturing, the ways we interact with computers are diversifying as never before [4].

Speech-recognition systems support a different kind of 'natural' interaction, allowing people to issue commands and dictate through voice. **Multi-touch surfaces** enable interaction with the hands and the fingertips on touch-sensitive surfaces, allowing people to manipulate objects digitally as if they were physical. Within **tangible interfaces** physical objects are embedded with computation, sensing and reacting

¹ For example, the first International Conference on Human-Robot Interaction was sponsored by ACM's Computer-Human Interaction Special Interest Group

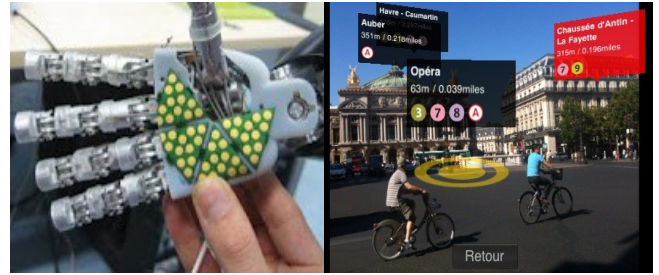


Figure 3. Left: Pieces of touch sensing skin designed for the humanoid robot iCub by Italian Institute of Technology in Genoa.

Right: The first augmented reality app called *Metro Paris Subway* showing bus and subway routes, locates nearby travel stations, and even highlights points of interest along the way, such as restaurants and hotels.

to the ways they are picked up, manipulated, and moved in space. The ability to sense our interaction without direct physical engagement with computer systems or input devices is also a growing trend. The visual based HCI is probably the most widespread area in HCI research. Researchers tried to tackle different aspects of human responses which can be recognized as a visual signal, e. g. Facial Expression Analysis, Body Movement Tracking, Gesture Recognition or Gaze Detection (Eyes Movement Tracking). **Eye movements** have been used for many years as a way of supporting the disabled in interacting with computers, but now we are also seeing the advent of '**brain computer interfaces**'. These systems can be useful for disable people using brain waves to interact with their environment. Real-time brainwave activity is beginning to be used to control digital movies, turn on music, and switch the lights on and off. The most difficult and costly devices to build are **haptic devices**. These kinds of interfaces generate sensations to the skin and muscles through touch, weight and relative rigidity.

Ambient Intelligence(AmI) is defined by [5] as „the convergence of ubiquitous computing, ubiquitous communication, and interfaces adapting to the user, “ where ubiquity involves the idea that something exists or is everywhere at the same time on a constant level, for example, hundreds of sensors placed throughout a household. The objective of AmI is to broaden the interaction between humans and digital information technology through the use of ubiquitous computing devices.

Virtual Reality replaces the real world with a simulated one, instead the **Augmented Reality** (following Wikipedia) s a live, direct or indirect, view of a physical, real-world environment whose elements are augmented by computer-generated sensory input. These paradigms are used especially in robotics research for testing the robot's performance.

B. HRI versus HCI

Authors in [7] identify several reasons why autonomous robots are a distinctive case of study than other technologies studied in HCI. They argue that people seem to **perceive autonomous robots differently** than they do most other computer technologies. People's mental models of autonomous robots are often more anthropomorphic than are their models of other systems – we expect more from robots than from conventional computers. Secondly, robots are ever more likely to be fully **mobile**, existing in the physical proximity with other robots, people, and objects. Mobile robots have to negotiate their

interactions in a dynamic, sometimes physically challenging, environment, creating a complex feedback system. Furthermore, robots **learn** about themselves and their world, and they exert at least some control over the information they process and actions they emit. An autonomous robotic system adds more complexity because it must adjust its decisions sensibly and safely to the robot's abilities and to the options available to the robot in a given environment. The system also must detect and respond to changes in the environment and its users.

Their [7] main argument is that a conventional computer looks like a machine and acts like one and robots are designed to resemble humans and mimicking people's behaviour. We argue to this opinion up to some point, as conventional computers are changing to new kinds of interfaces described in Section II. All of interfaces mentioned have application potential in robotics. For example, robots can learn in augmented reality and use this knowledge in „real“ reality. For example, such techniques are used to support remote interactions in NASA's Robonaut. Brain computer interfaces can control robot arms, allowing paralysed individuals to manipulate objects. Haptic devices are designed for humanoid robots (in Europe especially Italian Institute of Technology for humanoid iCub). Some suggest that telepresence, the natural extension of human awareness of a remote space, is a goal of interface design in HRI. Also, game industry is a wide application area, sophisticated multi-player online games may become useful in understanding how natural language can be used to support HRI and how human-robot teams should interact.

Physical embodiment via a robot opens up a number of new research challenges to us. Unlike virtual agents, robotic agent co-habit with the human in a physical space in which social interactions take place.

III. ROBOTS THAT PREDICT WHAT A PERSON IS WANTING

New trend in HRI is to program robots that respond and adapt to people's needs accordingly. In the past, most machine-learning applications operated 'off-line', where a set of training data would be collected and used to fit a statistical model. The fields of artificial intelligence (AI) and cognitive science have a great deal of relevance to the field of HRI. Cognitive models are being used for modelling how a human might interact and as the basis for generating robot behaviour.

As mentioned before, robots are expected to interact with humans in social environments, like hospitals, schools or at home, a comfortable and **intuitive way of communication** has to be established. Case study with platform IROS in Section IV.A illustrates such a communication. This platform has a system for identifying human intentions so it predicts his/her goals. In such manner it helps human in a collaborative task efficiently. When humans communicate with each other, they do not only use speech to convey the content of a message. At the same time, they employ a large variety of emotional and social signals to express consciously or unconsciously additional information, for example, about their attitude towards the conversational partner, their level of attention and their personality. In Section IV.B we present



Figure 4. Interaction with a platform IROS, research supported by EuCog II. The robot is endowed with a system for prediction of human's goals for a collaboration task.

a case study built on platform the NAO where a robot possesses an emotional model.

Researchers in different fields of developing technologies consider social and emotional levels of interaction as a critical role in a person's acceptance of and overall experience with any technology. Emotions are an essential part of human intelligence, and play a crucial role in perception, rational decision making and learning. By providing a robot with a personality, it helps provide people with good models and good understanding of the behaviour.

Marvin Minsky in his *The Society of Mind* [8] reasons about artificial emotions: "the question is not whether intelligent machines can have emotions, but whether machines can be intelligent without any emotions."

IV. CASE STUDIES WITH REAL ROBOTS

A. Case Study with IROS: Predicting Human's Goals

Based on the research of [9], we performed experiments with the robot IROS (Figure 4.), which has verbal and non-verbal communication skills for natural and efficient HRI. It is endowed with an architecture which implements a flexible mapping from observed action of a user onto "to-be-executed" complementary behaviour. This output behaviour is a speech output and/or a goal-directed action.

The mapping takes into account the inferred goal of the user (team-partner), shared task knowledge and contextual cues to accomplish a task. First of all, the robot has to infer what object the human user intends to build. Subsequently, the team formed by the robot and one human partner, constructs the target object from its components following the assembly plan. An action monitoring system which detects a mismatch between predicted and perceived action outcomes is also presented. Its direct link to the robot's motor representations of complementary behaviours guarantees the alignment of actions and decisions between the co-actors also in trials in which the human user performs unexpected behaviour.

His mental model for making decisions is based on the theory of dynamic neural fields, explained in [10]. The snapshots of video sequences, which can be found at <http://youtu.be/8ofnNRtliT8>, shall illustrate the processing mechanisms underlying the robot's capacity to anticipate the user's needs and to deal with unexpected events.

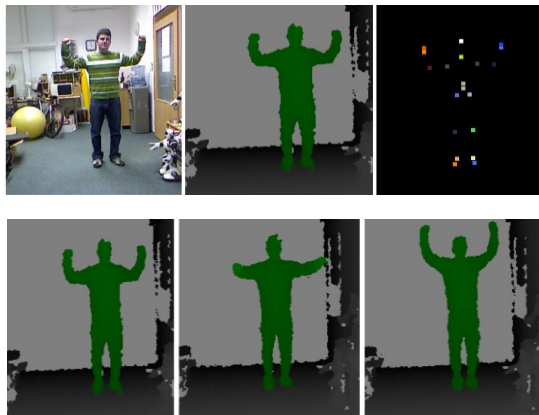


Figure 5. Up: Human expressing his emotional state. From left to right: RGB camera view, Kinect depth sensor view, skeleton extraction. Down: Human expressing JOY – a sequence in time

B. Case Study with NAO: Predicting User's Behaviour by Emotional Analysis

In recent years huge progress was made in the effort to express emotions with humanoids, mainly with facial expressions. Our results with the implementation based on the Body Movement expressions suggest that a humanoid robot, such as Nao, should be able to display emotions using its body (via postures, body movement and proxemics) Postures, specific positioning that the body takes during a timeframe, postures are an effective medium to express emotion, also many emotions are differentiated by characteristic body movements, and that these are effective cues for judging the emotional state of other people even in the absence of facial and vocal cues. Body movements include the movements themselves as well as the manner in which they are performed. Proxemics is the distance between individual during a social interaction, which is also indicative of emotional state.

To develop our model, we propose to use a Membership-function ARTMAP neural network (NN), a clustering technique with supervision, in order to learn different expressions of emotions from various users. Moreover, this paper seeks to demonstrate the usefulness of our methodology by adapting a model from psychology developed by R. Plutchik. It consists of eight basic emotional states and all other emotions are mixed or derivative states; that is, they occur as combinations, mixtures, or compounds of the primary emotions. Firstly, a database of expressions which describe these primary emotions is constructed, serving as input data for our NN. The system is able to recognize emotional expressions of various users and is transferred to a humanoid robot. Afterwards, the robot is capable to recognize emotional expressions of different users interacting with the robot and also can express its internal state by adopting the emotional expressions. Thus he can predict user's intentions based on the current emotional states. More about this research can be found in [11].

V. CONCLUSION

When the robot expresses his internal states via his non-verbal and verbal expressions during communication, humans should better understand him. The question is, if robot can express emotions understandable by human society, he should learn to recognize people's intentions to help him as a team partner in collaboration tasks.

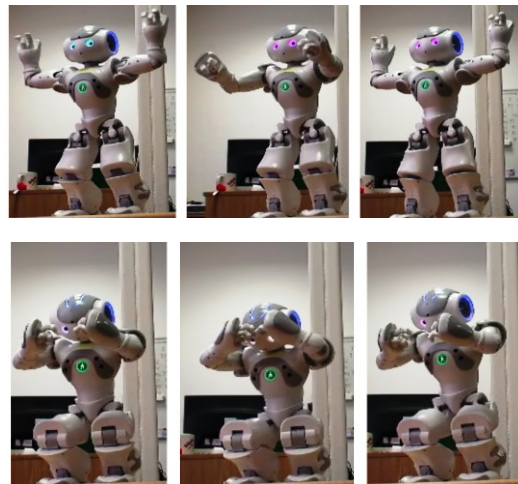


Figure 6. Nao expressing emotions learned from a User. Up: JOY. Down: FEAR. This is the second phase of the project, in which he learns to imitate user's emotional expressions. We suppose this way the user should better understand him.

Our future research will concentrate on the implementation of reactive social behaviours. NAO should be able to analyze social signals from the human during their collaboration and respond to it in real time. It is a complex task, including adaptation, learning, memory, motivation, focus of attention, internal model, communication and exchange of information, sensory integration, objectives management (creation of new goals, priorities) up to the emergent phenomena. Robotic sensing, cognitive, and actuating capabilities need to achieve a certain level of complexity such that people could treat them more as team-mates or companions to move them out of industry into everyday life.

ACKNOWLEDGMENT

Research supported by the "Center of Competence of knowledge technologies for product system innovation in industry and service", with ITMS project number: 26220220155 for years 2007 – 2013 and "Development of the Center of Information and Communication Technologies for Knowledge Systems" (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] AISB'05: "Social Intell. and Interaction in Animals, Robots and Agents", University of Hertfordshire, Proceedings of the Symposium on RobotCompanions: Hard Problems and OpenChallenges in HRI, 2005.
- [2] Human-Robot Interaction forum, <http://humanrobotinteraction.org/>
- [3] HRI Group at Kaist, South Korea, http://robot.kaist.ac.kr/?page_id=111
- [4] Microsoft Research Symposium on HCI, <http://research.microsoft.com/en-us/um/cambridge/projects/hci2020/downloads/beinghumanffa3.pdf>
- [5] M. Gupta, "Amb. Intell. - unobtrusive technology for the inf. society", in Pressbox.co.uk, 2003.
- [6] M.S. Raisinghani, A. Benoit, J. Ding, et.al. "Amb. Intell.: Changing forms of HCI and their Soc. Implic.," in Journ. Digital Inf. 5, No 4, 2004.
- [7] S. Kiesler, P. Hinds, "Introduction to this Special Issue on HRI," in CMU and Stanford Univ. Special Issue of HCI, 19, No 1, 2, 2004.
- [8] M. Minsky, "The Society of Mind," Simon and Schuster, NY, 1988.
- [9] E. Bicho, L. Louro, W. Erlhagen, "Integrating Verb. and Nonver. Comm. in a Dynamic Neural Field Arch. for HRI", in Front. Neurobot., 2010.
- [10] W. Erlhagen, E. Bicho, "The dynamic neural field approach to cognitive robotics," Journal of Neural Eng., vol. 3, 2006, pp. 36–54.
- [11] M. Vircikova, P. Smolar, P. Sincak, "Neural Approach for Personalised Emotional Model in HRI", to appear in IEEE World Congress on Computational Intelligence, Australia, 2012 (submitted).

Experiments on Language Patterns for New Approach in Formation of Programming Languages

¹Emília PIETRIKOVÁ (2nd year), ²Sergej CHODAREV (3rd year)

Supervisor: ³Ján KOLLÁR

^{1,2,3}Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹emilia.pietrikova@tuke.sk, ²sergej.chodarev@tuke.sk

Abstract—The aim of this article is to present results of experiments performed on a large group of Haskell programs, providing sample derivation trees with recognized language patterns and statistics as table of the most frequent language symbols. These experiments lead towards new conception and features of language patterns, with the rise of language expression ability, covering current paradigms. The main contribution lays in new approach within development of programming languages, language dialects and domain-specific constructions.

Keywords—Derivation tree, programming languages, syntactic patterns, domain-specific languages.

I. INTRODUCTION

To propose a solution for automatized introduction of new language abstractions that are based on patterns found in source code, problem of recurring pattern recognition should have been solved. Manual analysis of the code may be hard and tedious task. However, a tool for automatic pattern recognition can greatly help in this task. Moreover, the recognition needs to be done at the level of program syntax.

The term of program pattern means code fragment extracted from a set of sample programs that have equivalent syntactic, and hence, also semantic structure. Patterns can also contain parts that are different in each program. These parts can be called syntactic variables.

Expressiveness of a language can be improved by the recognition of program patterns. Moreover, it allows more natural and strain-forward expression of programs. This approach can also be useful for development of domain-specific dialects of programming languages. In order to implement this transition from general-purpose language (GPL) to its domain-specific dialect, it is necessary to reflect the fundamental differences between the domain-specific dialect and the corresponding GPL. The main differences lie in the following:

- focus on a particular domain,
- use of concepts from a domain,
- higher abstraction.

To achieve a connection with particular domain and a shift towards domain specificity, it is suitable to analyze existing programs (or program fragments) written in the GPL solving various problems from the domain. On the basis of this analysis, a shift from GPL to domain-specific dialect can be achieved, overcoming the mentioned differences as follows:

- *Domain specificity* — DSL is aimed at solving problems of a particular domain and consists of structures and notation associated with the domain. Thus, it is possible

to identify linguistic structures that are not used in programs addressing problems of the domain.

- *Using concepts from the domain* — DSL uses concepts of a problem domain and defines relations among them. Thus, it is essential to find and identify domain-specific constructs that are repetitive in particular programs.
- *Higher abstraction* — GPLs are intended to solve various problems, consequently they contain only general implementations and abstraction of lower levels. They are used in order to create solution to a specific problem. On the other hand, DSLs are dedicated for particular domains, thus containing specific solutions and implementations in the form of a higher level of abstraction. Therefore during the analysis, patterns recurring in individual programs were searched, so it was possible to unify and create higher level abstractions.

In pursuance of these facts, implementation of a domain-specific dialect from the base language consists of two parts:

- a) introducing new syntactic elements for abstractions used in the domain — *language extension*,
- b) removing syntactic elements not used/needed in programs for the domain — *language reduction*.

Reduction of the base language is very important and should not be overlooked. By reducing unneeded syntactic elements, the language becomes easier to learn. It also decreases possibility of errors that may result from accidental usage of wrong language elements. Moreover, reduction of unneeded elements can also allow syntax simplification of the rest of the language.

II. METHODOLOGY OF EXPERIMENTS

For experimental purposes, Haskell 98 was chosen as a language for the analysis. To get a proper knowledge about language constructs and syntactic structure of the analyzed programs, a complex set of tools has been developed. As a result of one program analysis, derivation tree is produced, consisting of the used Haskell grammar rules [1].

Architecture of syntax analysis consists of two parts — *generating* and *analyzing infrastructure*. The goal of generating infrastructure is to prepare tools being used during the analysis, and the analyzing infrastructure contains lexical analyser (lexer) and parser, intended for analysis of specific programs into lexical units, then processing them into derivation trees. Derivation trees have been produced for further process to retrieve statistical data on the programs, and to recognize common language patterns.

Grammar Ambiguity

Meanwhile the parsing, several ambiguities were detected. It was necessary to modify grammar rules `pat_i` and `exp_i` to avoid a situation when the parser was not able to join parallel parsings.

It was also spotted in the grammar rules `export`, `import`, `aexp`, and `alts`, where parser could not determine, which alternative should be chosen. The problem was solved by setting priority using the `%dprec` directive.

Another ambiguity was related to `lambda (\)` and conditional (`if-then-else`) expressions. In these cases, parser could not determine correctly where the expression ended. According to *Haskell Report* [1], it is supposed to continue as far as possible. As a solution, the `%merge` directive was used with a custom merge function. With two possible subtrees given as arguments, the function decided which subtree was the correct one.

Fixity Resolution

Another problem arised with a resolution of fixity and precedence of operators. Instead of defining separate grammar rules for all precedence levels of expressions (like it is done in *Haskell 98 Report* [1]), infix expressions were defined using one common rule; fixity and precedence were resolved after parsing. Resolution was done according to the algorithm specified in *Haskell 2010 Report* [2].

III. CODE STATISTICS

Using developed tools, it was possible to compute some interesting statistics based on a set of about 300 Haskell sample programs. As a result of program analysis, its derivation tree is provided according to the language grammar. The derivation tree consists of terminal and nonterminal symbols, where terminal symbols represent leaves of the tree. The derivation tree also contains helper nodes corresponding to EBNF features like repetition or optional elements.

One of the parameters that may be investigated, is a relative occurrence of symbols in derivation trees. Relative occurrence of a symbol in a program is defined as $r_{sym} = \frac{n_{sym}}{N}$, where n_{sym} means a number of occurrences of the `sym` symbol in the derivation tree and N represents a number of all symbols/nodes of the derivation tree.

Table I represents 40 most frequent occurrences of particular symbols in all programs of our sample. As it can be expected, variable names and expressions have the greatest frequency. However, some symbols even did not occur in any of our sample programs, like `default`, `fbind`, `fpat` and `gdp`.

It is possible to provide similar statistics for specially selected sample of programs within a specific domain. This might show, which language elements are used in programs of the domain and which elements can be omitted from the domain-specific dialect. Moreover, statistical analysis can also be used to partition a sample of programs into groups based on a usage of the language elements.

IV. PATTERN RECOGNITION

To recognize syntactic patterns in a program or a set of programs, it is important to decide which parts of analyzed programs may be considered similar. The simplest possibility is to consider only the equal trees. However, this approach

TABLE I: Proportion number of 40 most frequent symbol occurrences

Symbol	Occurrence	Symbol	Occurrence
varid	0,093855	qvarop	0,015110
aexp	0,092660	gcon	0,014879
fexp	0,092660	atype	0,013402
exp_10	0,063059	varsym	0,012450
qvar	0,051154	qcon	0,011951
exp_i	0,049428	btype	0,011516
exp	0,044523	,	0,011309
var	0,037632	funlhs	0,010876
apat	0,033349	type	0,010080
conid	0,026202]	0,008857
(0,019259	[0,008857
)	0,019259	integer	0,008808
=	0,018341	gtycon	0,007696
decl	0,017526	pat	0,007687
:	0,017277	string	0,005727
qop	0,016620	-i	0,005454
topdecl	0,016484	}	0,004845
rhs	0,016164	{	0,004845
pat_i	0,016099	—	0,004296
pat_10	0,016099	qconop	0,003216

is exceedingly limiting. Trees can be considered similar if their structure is the same except for the attributes of terminal symbols (approach that has been chosen).

Another approach is to allow differences in whole subtrees rooted in a node of the same type. This would allow more complex syntactic variables, but it is harder to implement.

To find patterns in the program derivation tree, a simple algorithm can be used that is based on the function *findPatterns* defined below:

```

parents ← allParents(elements)
groups ← findGroups(parents)
if groups is empty then
  return [groups]
else
  for all group ∈ groups do
    Add findPatterns(group) to foundGroups
  end for
  return mergeGroups(foundGroups)
end if

```

Function *findPatterns* takes a list of the tree elements and recursively examines their parents to find a set of groups of subtrees that have a similar structure. It uses helper functions with the following meaning:

- *allParents* — returns a set of parents of all tree elements in a group;
- *findGroups* — given a set of tree elements, returns list of groups of elements with similar subtrees;
- *mergeGroups* — merges list of group lists into a single list.

To initiate the algorithm, the *findPatterns* function is called on terminal symbols of the tree. Then it tries to walk up to the root of the tree while it can find groups of subtrees with similar structure. List of subtree groups is a result of the algorithm, where each group corresponds to a found pattern and contains all occurrences of the pattern.

Example

Lets look at a simple example program. It defines function *eval* evaluating expressions defined using derived abstract syntax tree. Derivation tree of this program is represented in Fig. 1.

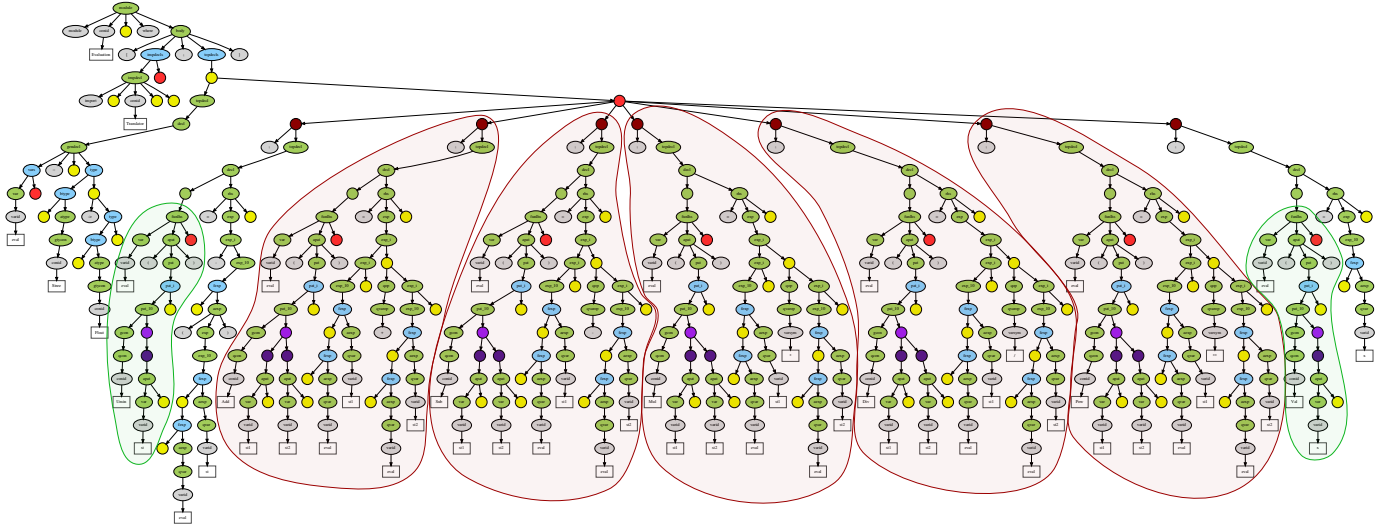


Fig. 1: Derivation tree of the example program with recognized patterns.

```
module Evaluation where
import Translator
```

```
eval :: Stree -> Float
eval (Umin st)      = -(eval st)
eval (Add st1 st2)  = eval st1 + eval st2
eval (Sub st1 st2)  = eval st1 - eval st2
eval (Mul st1 st2)  = eval st1 * eval st2
eval (Div st1 st2)  = eval st1 / eval st2
eval (Pow st1 st2)  = eval st1 ** eval st2
eval (Val x)        = x
```

Using the described method, it is possible to find several recurring patterns in this program (see Fig. 1). The most important are:

- $\text{eval } (\alpha \text{ st1 st2}) = \text{eval st1 } \beta \text{ eval st2}$
- $\text{eval } (\alpha \beta)$

Greek letters in the patterns represent syntactic variables that can be replaced with concrete syntactic elements. Other found patterns are too small to be mentioned.

V. CONCLUSIONS AND FURTHER RESEARCH

In this article, we presented experiments that made it possible to accomplish pattern recognition in program codes, with the perspective of development of new language dialects, both general-purpose and domain-specific. The term of program patterns has been used for syntactically, and hence, also semantically equal program fragments occurring in a set of program samples. As it has been shown, having defined grammar of a language, as well as a set of program samples, we are able to evaluate the frequency of the use of symbols — concepts in the language.

This may be interesting from the perspective of language benchmarking, the goal of which is to reduce the amount of redundant constructs. Thus, further research also involves an extension for processing a whole set of programs. Another usage might be extension of a language based on the needs of programmers [3]. It may allow adding new constructs to the language, that correspond to repeated code fragments.

However, upon the presented results, the most significant is contribution for automated software evolution. Clearly, this would mean to shift from a language analyzer to language

abstracter, associating concepts to formal language constructs [4], and formalizing them by means of these associations. In this way, we expect to integrate programming and modeling, associating general purpose and domain-specific languages [5], [6], as well as to perform qualitative movement from an automatic roundtrip engineering [7], [8] to the automated roundtrip software evolution, that is understood as the software development without any affects of a human.

Therefore, experiments performed in this article outcome additional experiments with new conception and features of language patterns, meaning the rise of language expression ability, covering current paradigms. Next, the research will focus on new methods of composition (or combination) of programs, based on the new conception of language patterns. This new approach would also mean a theoretical contribution to language grammars based on new pattern conception, with the aim of flexibility increase within specialization of the patterns in particular fields of application.

ACKNOWLEDGMENT

The work presented in this paper was supported by SRDA Project of Slovak-Slovenian Research and Development No. SK-SI-0003-10 "Language Patterns in Domain-specific Languages Evolution".

REFERENCES

- [1] S. Peyton Jones, *Haskell 98 Language and Libraries – The Revised Report*. Cambridge, England: Cambridge University Press, 2003.
- [2] S. Marlow, "The Haskell 2010 Language Report," Available: <http://www.haskell.org/onlinereport/haskell2010/>, 2010.
- [3] G. L. Steele, "Growing a language," *Higher-Order and Symbolic Computation*, vol. 12, pp. 221–236, 1999.
- [4] J. Porubán and P. Václavík, "Extensible language independent source code refactoring," in *AEI '2008: International Conference on Applied Electrical Engineering and Informatics, Greece, Athens, September 8-11*. Košice: FEI TU, 2008, pp. 58–63.
- [5] M. Sabo and J. Porubán, "Preserving design patterns using source code annotations," *Journal of Computer Science and Control Systems*, vol. 2, no. 1, pp. 53–56, 2009.
- [6] I. Luković, P. Mogin, J. Pavićević, and S. Ristić, "An approach to developing complex database schemas using form types," *Software Practice & Experience*, vol. 37, no. 15, pp. 1621–1656, 2007.
- [7] U. Almann, "Automatic roundtrip engineering," *Electronic Notes in Theoretical Computer Science*, vol. 82, no. 5, pp. 33–41, 2003.
- [8] C. Lohmann, J. Greenyer, J. Jiang, and T. Systä, "Applying triple graph grammars for pattern-based workflow model transformations," *Journal of Object Technology*, vol. 6, no. 9, pp. 253–273, 2007.

Explicit and Online Predictive Control of Nonlinear Hydraulic System

¹Štefan JAJČIŠIN (2nd year)
Supervisor: ²Anna JADLOVSKÁ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹stefan.jajcisin@tuke.sk, ²anna.jadlovaska@tuke.sk

Abstract—An explicit and online predictive control of nonlinear hydraulic system is presented in this paper. Obtained results of explicit predictive control are compared with results by the classical approach of predictive control, where an optimization task is executed in every sample of control closed loop. According to particular introduced algorithms program modules in simulation language Matlab are created and used in the simulation control of nonlinear hydraulic system. The main goal of this paper is to compare both approaches of predictive control of nonlinear dynamical systems based on a predictor in a linear form, summarize advantages and disadvantages of each approach.

Keywords—nonlinear hydraulic system, state-space model based predictive control, explicit predictive control.

I. INTRODUCTION

The paper deals with the predictive control of nonlinear hydraulic system. Predictive control algorithms, which were used in the control are based on using a linear approximation of nonlinear model of controlled physical system. As controlled system the nonlinear simulation model of hydraulic system, which is located in the Laboratory of mechatronic systems at the Department of Cybernetics and Artificial Intelligence was used. The simulation of its control was carried out in Matlab/Simulink with created program modules in the simulation language of Matlab on the basis of theoretical background of predictive control algorithms.

As the hardware configuration and the mathematically-physical description of used hydraulic system were presented in [1], in this paper it is introduced very briefly in the part II. Next the basic principle of predictive control on the basis of predictor in a linear form, the theoretical derivation and the programming design of control algorithms are mentioned. The part III is devoted to predictive control algorithms, which is based on the state space model of controlled system. In the part IV the explicit solution of predictive control is presented. In the end of this paper results of nonlinear hydraulic system simulation control by mentioned algorithms are depicted and mutually compared.

II. NONLINEAR HYDRAULIC SYSTEM

As it have already been mentioned in the introduction, the

used nonlinear model of hydraulic system was introduced in [1], eventually in [2], where its principle of dynamics, the hardware configuration, the communication way and the control by digital PID algorithms running in Matlab on PC level were alledged. For that reason only physical structure and a systemic view of this model for control purpose is presented in this paper. A schematic illustration of 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).

Physical quantities shown in Fig. 1 are:

- $f_m(t)$ - 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} .

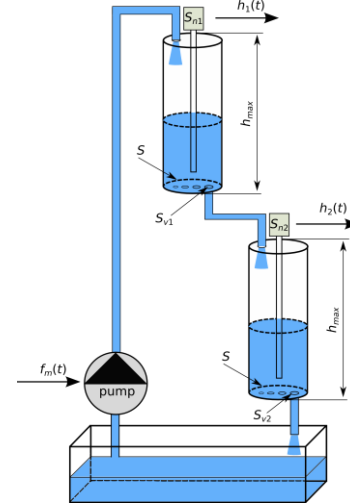


Fig. 1: The hydraulic system of two tanks

The systemic view of introduced hydraulic system is depicted in Fig.2, where besides already mentioned quantities, $q_{in1}(t)$ and $q_{in2}(t)$ is inflow to the first and the second tank.

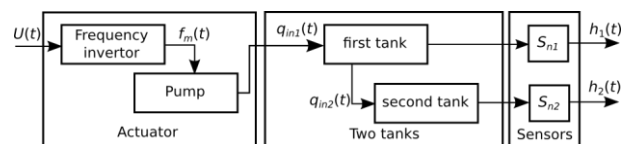


Fig. 2: The systemic view of hydraulic system

Regarding to the control process, a voltage of frequency inverter in range 0 – 10V constitutes the control action quantity and the liquid level in the second tank $h_2(t)$ is the controlled quantity. State quantities are liquid levels in both tanks $h_1(t)$ and $h_2(t)$, which have limit values 0 – 0.3m.

III. ONLINE PREDICTIVE CONTROL

The basic principle of predictive control algorithms consists in using linear model to compute a prediction of future behaviour of controlled system on the length of prediction horizon (Fig. 3). Regarding to used linear model it is possible to divide the set of predictive control algorithms into two categories:

- A. algorithms of **Generalized Predictive Control (GPC)**, which use input/output description [3],
- B. **state Space Model-based Predictive Control** algorithms (SMPC).

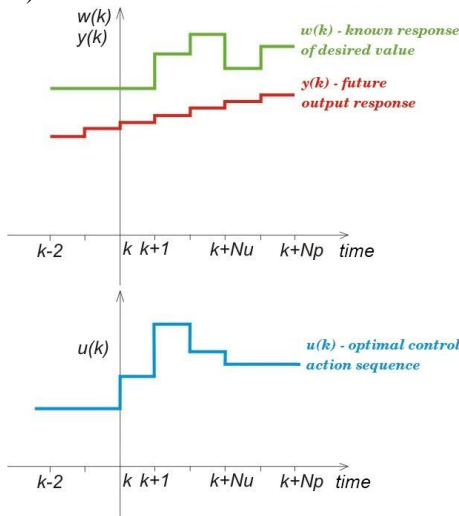


Fig. 3: Predictive control principle

The general control structure with predictive control algorithm is depicted in Fig. 3, where w is vector of output desired on the length of prediction horizon, $u(k)$ is control action computed by GPC or SMPC algorithm, $d(k)$ is disturbance vector and $y(k)$, $x(k)$ is output, state vector of controlled system, respectively.

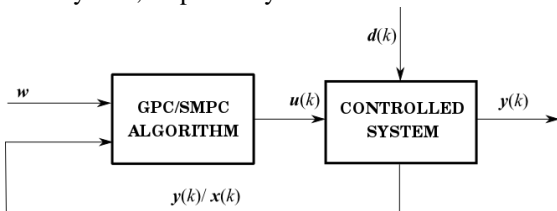


Fig. 4: The control structure with GPC/SMPC algorithm

Control action computing is based on a criteria function minimization J_{MPC} , where setting parameters are represented by weighing matrices Q and R . Thus, it is carried out by optimization task. Additionally, predictive control algorithms have a big advantage, that it is possible to transform control action computing to the quadratic programming task and involve physical quantities constraints of controlled system. The criteria function minimization with constraints by quadratic programming means more computing demands, because it is possible to carry out such calculation only by numerical way.

Control action computing by predictive control algorithms is carried out on the basis of receding horizon strategy. The optimization task is executed at each sample period and results to the optimal sequence of control action. However, only the first element of computed sequence is used as system control input. This procedure is repeated again in next sample instances.

Relatively, many modifications of basic predictive control principle exist. In this paper two algorithms of SMPC are mentioned.

In this part an online predictive control algorithm, which keeps the receding horizon principle is introduced. It uses a discrete state space description of dynamic system

$$\begin{aligned} x(k+1) &= A_d x(k) + B_d u(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned} \quad (8)$$

where $x(k)$, $u(k)$ and $y(k)$ is a vector of system's states, inputs and outputs, and matrices A_d , B_d , C , D with particular dimensions contain coefficients represented system dynamics for predictor derivation.

According to [5] in the case of SMPC algorithm it is used criteria function

$$\begin{aligned} J_{MPC} &= \sum_{i=N_1}^{N_p} Q [\hat{y}(k+i) - w(k+i)]^2 + \\ &+ \sum_{i=1}^{N_u} R [\Delta u(k+i-1)]^2 \end{aligned} \quad (9)$$

where N_p and N_u is prediction and control horizon, Q and R are weighing matrices and $\hat{y}(k)$ is predicted output.

Regarding to the fact, that we want to weigh control action rate $\Delta u(k)$ in the criteria function, we can isolate $\Delta u(k)$ in predictor derivation according to [5]:

$$\begin{aligned} x(k+1) &= A_d x(k) + B_d u(k-1) + B_d \Delta u(k) \\ y(k) &= Cx(k) + Du(k) \\ u(k) &= u(k-1) + \Delta u(k) \end{aligned} \quad (10)$$

According to [6] and provided that $D=0$ it is possible to derive the predictor by iteration of discrete state space description equations (10) step by step in the form

$$\hat{y} = Vx(k) + G_1 u(k-1) + G_2 \Delta u = y_0 + G_2 \Delta u, \quad (11)$$

where

$$\begin{aligned} G_1 &= \begin{pmatrix} CB_d \\ C(A_d + I)B_d \\ \vdots \\ C(A_d^{N_p-1} + \dots + A_d + I)B_d \end{pmatrix} \\ G_2 &= \begin{pmatrix} CB_d & 0 & \dots & 0 \\ C(A_d^{N_p} + I)B_d & CB_d & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ C(A_d^{N_p-1} + \dots + A_d + I)B_d & \dots & C(A_d + I)B_d & CB_d \end{pmatrix} \end{aligned} \quad (12)$$

After the predictor equation (11) substitution to the matrix form of criteria function J_{MPC} (9)

$$J_{MPC} = (\hat{y} - w)^T Q (\hat{y} - w) + \Delta u^T R \Delta u \quad (13)$$

and after multiplying it is possible to express it by the quadratic form

$$J_{MPC} = c + 2g^T \Delta u + \Delta u^T H \Delta u, \quad (14)$$

where g^T is gradient, H is Hessian and c is a constant.

According to [4] an analytic formula for optimal control action on the length of control action N_u

$$\Delta u = -H^{-1}g \quad (15)$$

can be expressed on the basis of condition

$$\frac{\partial J_{GPC}}{\partial \Delta u} = 0. \quad (16)$$

In the case when constraints of physical quantities regarding is required in control action computing, it is necessary to solve minimization of (14) by quadratic programming, for example with the *quadprog* function, which is part of *Optimization Toolbox* in Matlab.

The SMPC algorithm uses information about states $x(k)$ of controlled system in the feedback branch of control structure. The block of control algorithm in Fig. 4 can be shown as Fig. 5, where the prediction of system free response y_f is computed from current values of states $x(k)$.

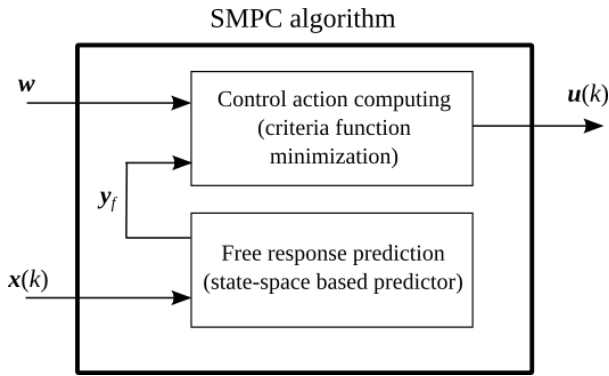


Fig. 5: The detail of SMPC algorithm block

IV. EXPLICIT PREDICTIVE CONTROL

The main disadvantage of predictive control algorithms, which are based on the receding horizon strategy is high calculation demands to criteria function minimization with constraints regarding. This drawback can be eliminated by an explicit solution of optimal control action computing in advance, before control process. As the first, the scientific group led by prof. Morari from ETH in Zürich was concerned in the explicit solution of optimal control with constraints [7]. Besides many things, results of their research is the Multiparametric toolbox (MPT), which contains functions for design, analysis and simulation of dynamic systems control by explicit predictive control on the basis of multiparametric programming [8].

According to [9] the result of multiparametric quadratic programming is explicitly computed control action $u^*(k) = f(x(k), w(k))$ for possible values of states $x(k)$, which can be used in the control structure in consequence.

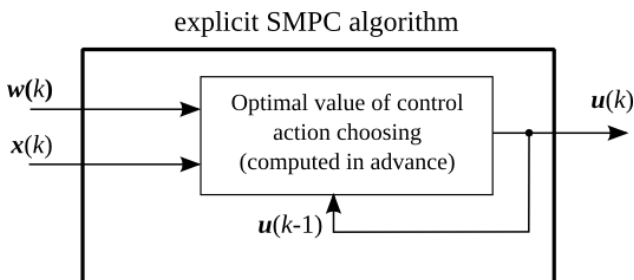


Fig. 6: The detail of explicit SMPC algorithm block

Forasmuch as the control algorithm selects an appropriate control action depending up to the current and desired state of controlled system in the frame of control process cycle, it is possible to use the structure depicted in Fig. 3 for dynamic system control. The detail of explicit predictive control algorithm block structure is shown in Fig. 6.

The basic algorithm, which makes possible to create a program of control structure for dynamic systems simulation control with MPT toolbox's functions is depicted in Fig. 8.

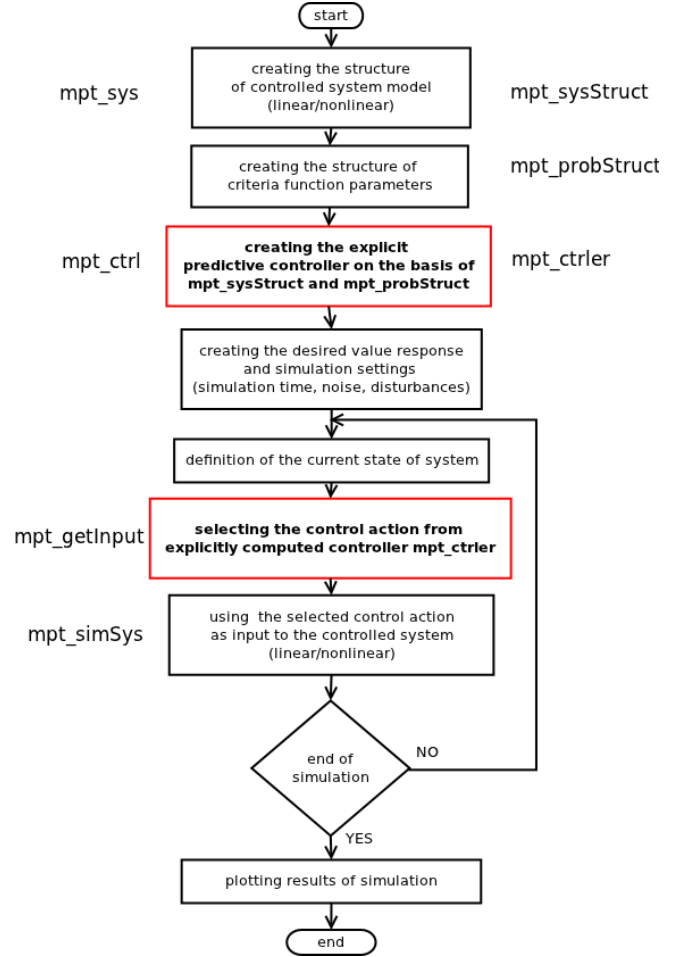


Fig. 7: The flow chart for programming the control structure with explicit predictive control algorithm

V. CONTROL RESULTS COMPARISON

In this part results of nonlinear hydraulic system control with both mentioned algorithms are shown, whereby we used next settings of criteria function parameters: $N_p = 10$, $N_u = 2$, $Q = 100I$, $R = 0.001I$, sample period $T_{vz} = 1s$. In Fig. 8 results of control, where the desired value of liquid level in the second tank had constant value 0,13m is depicted. Because of fast liquid level onset in the first tank it was necessary to limit the input voltage to 8V, otherwise a liquid overflow in the first tank would happen. In Fig. 9 results of control with variable desired value are shown. While results of control to constant value are similar, in control, when the desired value is varying, a late reaction of the explicit predictive control algorithm is apparently noticeable. This kind of shift is caused by the fact that the MPT predictive control algorithm does not take the consideration to desired value time response on the length of prediction horizon, but only in the current sample instance.

Regarding to the speed of control action computation in each control process sample instance the explicit solution was almost three times faster then the online approach. An obtained average spending time of one control process cycle in simulation control on the same computer (CPU Intel i5-2410M 2,3GHz, RAM 4GB, Win7 64-bit) was: online – 5.2ms, explicit – 1.8s.

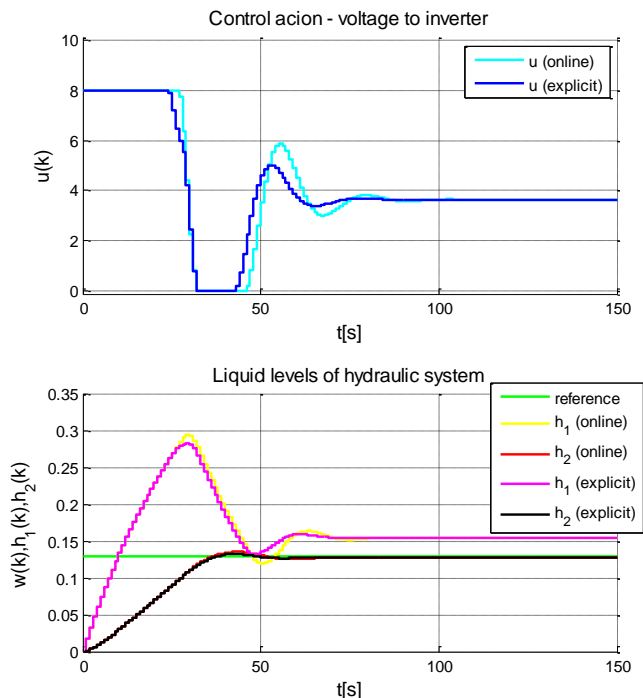


Fig. 8: Results of online and explicit predictive control of hydraulic system – constant desired value

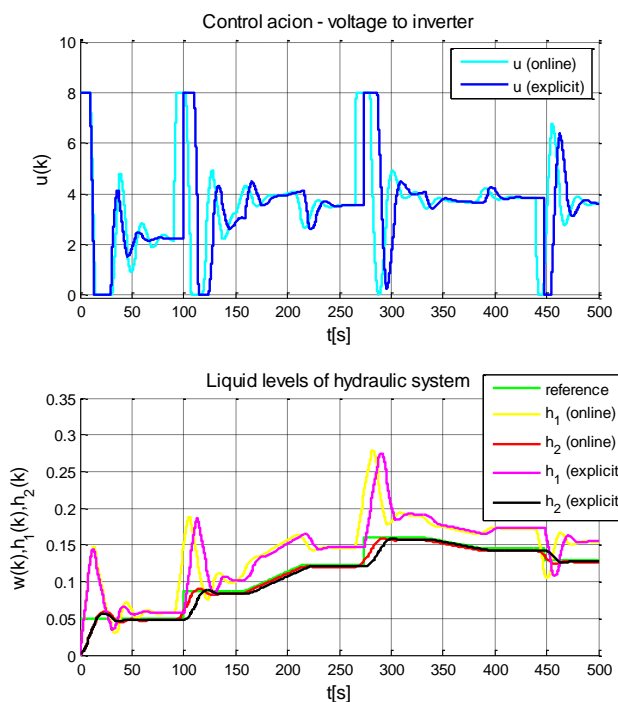


Fig. 9: Results of online and explicit predictive control of hydraulic system – variable desired value

VI. CONCLUSION

It results from time responses depicted in the previous part

that in the case when variable desired value is used it will be more preferable to use the online predictive control algorithm if fast achievement of desired value is required. However, in this case the computation demands of control algorithms should be taken into consideration. It is also necessary to adjust control requirements, for instance select suitable length of horizons, eventually pull out the sample time.

On the other side, in the case of physical system with fast dynamics control, when it is needed to use very short sample time, the explicit predictive control algorithm is more suitable than the online approach.

Both approaches of optimal control action computation are based on using the linear model of controlled physical system. However, using it in the explicit approach is more critical, especially when unmeasured disturbances may appear or in the case of control of systems, which dynamics is not possible to suitably approximate by linear model, whether for one or more operating points. This disadvantage is partly eliminated by control of piecewise systems, which is implemented enough in the frame of MPT toolbox, too.

ACKNOWLEDGMENT

This work has been supported by the Scientific Grant Agency of Slovak Republic under project Vega No.1/0286/11 Dynamic Hybrid Architectures of the Multiagent Network Control Systems (50%). This work is also the result of the project implementation Development of the Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (50%).

REFERENCES

- [1] Š. Jajčišin, "Verification of Control Algorithms with DDE Communication on Real Hydraulic System", in: SCYR 2011 : 11th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice : proc. - Košice : FEI TU, 2011 S. 388-391. - ISBN 978-80-553-0644-5.
- [2] Š. Jajčišin, A. Jadlovska, "Laboratory model of Hydraulic System Control" (in Slovak), in: Electroscope – online journal for Electrotechnics, year 2011, No. III. ISSN 1802-4564, Available on the internet: http://147.228.94.30/index.php?option=com_content&view=article&id=280:riadenie-laboratorneho-modelu-hydraulickeho-systemu-&catid=34:cislo-32011-&Itemid=48.
- [3] D.W. Clarke, C. Mohdani, P.S. Tuffs, "Generalized Predictive Control, Part 1 and 2", in Automatica, Vol 23, No.2, 137 – 160, 1987.
- [4] M. Fikar, "Predictive Control – An Introduction", Bratislava: Slovenská technická univerzita - FCHPT, 1999.
- [5] K. Belda, J. Böhm, M. Valášek, "Model-based control for parallel robot kinematics", Proceedings of the 3rd International Congress on Mechatronics. MECH2K4, p. 1-15, 2004.
- [6] J. Roubal, "Predictive controller", (in Czech), (Examples for exercises – Modern Control Theory). Available on the internet: http://support.dce.felk.cvut.cz/pub/roubalj/teaching/MTR/seminars/MTR_cv8_mpc.pdf
- [7] A. Bemporad, M. Morari, V. Dua, "The Explicit Linear Quadratic Regulator for Constrained Systems", Technical report, ETH Zürich, 1999.
- [8] M. Kvasnica, M. Hecceg, L. Čirka, M. Fikar, "Adaptive Model Predictive Control of Piecewise Affine Systems", in International Workshop on Assessment and Future Directions of Nonlinear Model Predictive Control, Pavia, Italy, 2008.
- [9] J. Řehoř, "Explicit Solving of Predictive Control Task", (in Czech) Diploma thesis, Praha: ČVUT-FE, 2008.

Swing-up and Stabilizing Control of Classical and Rotary Inverted Pendulum Systems

¹Slávka JADLOVSKÁ (1st year)

Supervisor: ²Ján SARNOVSKÝ

^{1,2}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹slavka.jadlovaska@tuke.sk, ²jan.sarnovsky@tuke.sk

Abstract— The purpose of this paper is to present the design of a complete control strategy for the classical and rotary single inverted pendulum system. The design, which involves swing-up and stabilization of both systems, is based on function blocks and GUI tools from the *Inverted Pendula Modeling and Control*, a *Simulink* block library developed by the author of the paper.

Keywords—classical single inverted pendulum, rotary single inverted pendulum, energy-based swing-up methods, state-feedback stabilization, custom *Simulink* block library

I. INTRODUCTION

Inverted pendula systems (IPS) represent a significant class of nonlinear underactuated mechanical systems, well-suited for the verification and practice of ideas emerging in control theory and robotics. Stabilization of a pendulum rod in the unstable upright position is considered a benchmark control problem which has been solved by attaching the pendulum to a base that either moves in a controlled linear manner (*classical IPS*) or in a rotary manner in a horizontal plane (*rotary IPS*).

As an example of task-oriented control of IPS, swinging the pendulum up from the pendant to upright position was introduced in [1]. Any system of inverted pendula is therefore a suitable testbed for illustrating hybrid control approaches.

The *Inverted Pendula Modeling and Control (IPMaC)* is a structured *Simulink* block library which was developed by the author of this paper and provides complex software support for the analysis and control of both classical and rotary IPS [2]. Strong emphasis is placed on the generalized approach to system modeling [3], allowing the library to handle systems which differ by the number of pendulum links attached to the base, such as single [2][3], double [2][4] and triple IPS.

This paper aims to present a control strategy which ensures successful swinging up and stabilization of classical and rotary single IPS. Each step of the design relies on suitable function blocks or GUI tools from the *IPMaC*. Since the stabilization problem has already been thoroughly dealt with (see [2][3][4]), the paper will focus on the swing-up design, notably on the comparison of several proposed methods.

II. MATHEMATICAL MODELING OF SINGLE INVERTED PENDULUM SYSTEMS – AUTOMATIC APPROACH

The *classical single inverted pendulum system* is composed of a rigid, homogenous pendulum rod hinged to a stable mechanism (cart) [5] which allows for free movement along

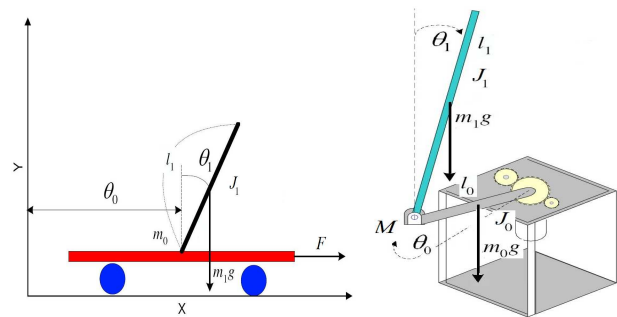


Fig. 1 Scheme and nomenclature: a) Classical single inverted pendulum system b) Rotary single inverted pendulum system

a single axis (Fig. 1a). The *rotary single inverted pendulum system* (i.e. the *Furuta pendulum* [1]) consists of a pendulum rod attached to an arm which is free to rotate in a horizontal plane (Fig. 1b) [6]. Both systems are underactuated: the only input (force applied on the cart or torque applied on the arm) is in each case used to control the two degrees of freedom of the system: base position $\theta_0(t)$ (cart position [m] or arm angle [rad]) and pendulum angle $\theta_1(t)$ [rad].

A. Modeling and Simulation of Considered Systems

The *Inverted Pendula Model Equation Derivator* (Fig. 2) is a MATLAB GUI application and a central component of the *IPMaC*. The application generates the motion equations for the user-chosen types of IPS (classical/rotary, single/double) using an original procedure of mathematical model derivation for the generalized (*n-link*) IPS, which was implemented in MATLAB using *Symbolic Math Toolbox* [3].

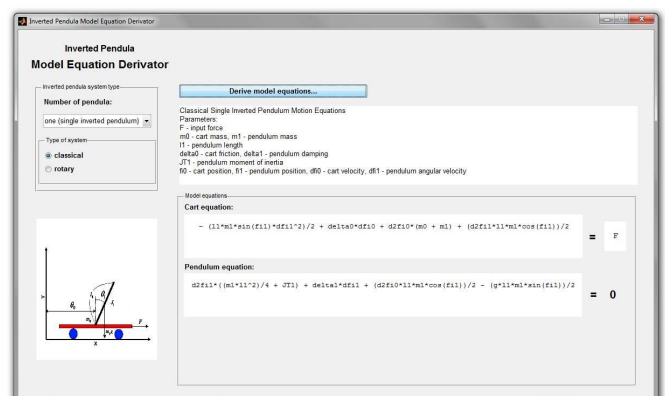


Fig. 2 *Inverted Pendula Model Equation Derivator*

With aid of the *Derivator*, mathematical models of both considered systems were obtained and will be presented below in the rearranged, so-called *standard minimal ODE (ordinary differential equation)* matrix form:

$$\mathbf{M}(\boldsymbol{\theta}(t))\ddot{\boldsymbol{\theta}}(t) + \mathbf{N}(\boldsymbol{\theta}(t), \dot{\boldsymbol{\theta}}(t))\dot{\boldsymbol{\theta}}(t) + \mathbf{P}(\boldsymbol{\theta}(t)) = \mathbf{V}(t) \quad (1)$$

where $\boldsymbol{\theta}(t) = (\theta_0(t) \ \theta_1(t))^T$. This approach allows to isolate $\ddot{\boldsymbol{\theta}}(t)$ and express both systems in the nonlinear state-space form

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t), u(t), t), \\ \mathbf{y}(t) &= \mathbf{g}(\mathbf{x}(t), u(t), t) \end{aligned} \quad (2)$$

by defining the state vector as $\mathbf{x}(t) = (\boldsymbol{\theta}(t) \ \dot{\boldsymbol{\theta}}(t))^T$.

After the rearrangement, the two second-order nonlinear differential equations of the *classical single inverted pendulum system*, which respectively correspond to the cart and the pendulum, assumed the form:

$$\begin{pmatrix} m_0 + m_1 & \frac{1}{2}m_1l_1\cos\theta_1(t) \\ \frac{1}{2}m_1l_1\cos\theta_1(t) & J_1 \end{pmatrix} \begin{pmatrix} \ddot{\theta}_0(t) \\ \ddot{\theta}_1(t) \end{pmatrix} + \begin{pmatrix} \delta_0 & -\frac{1}{2}m_1l_1\dot{\theta}_1(t)\sin\theta_1(t) \\ 0 & \delta_1 \end{pmatrix} \begin{pmatrix} \dot{\theta}_0(t) \\ \dot{\theta}_1(t) \end{pmatrix} + \begin{pmatrix} 0 \\ -\frac{1}{2}m_1g l_1 \sin\theta_1(t) \end{pmatrix} = \begin{pmatrix} F(t) \\ 0 \end{pmatrix} \quad (3)$$

where m_0 is the cart mass, m_1 is the pendulum mass, l_1 is the pendulum length, δ_0 is the friction coefficient of the cart against the rail, δ_1 is the damping constant in the joint of the pendulum, $J_1 = \frac{1}{3}m_1l_1^2$ is the pendulum's moment of inertia with respect to the pivot point and $F(t)$ is the force induced on the cart.

Analogically, the mathematical model of the *rotary inverted pendulum system*, composed of two second-order nonlinear differential equations which respectively describe the rotary arm and the pendulum, became:

$$\begin{pmatrix} J_0 + m_1l_0^2 + \frac{1}{4}m_1l_1^2\sin^2\theta_1(t) & \frac{1}{2}m_1l_0l_1\cos\theta_1(t) \\ \frac{1}{2}m_1l_0l_1\cos\theta_1(t) & J_1 \end{pmatrix} \begin{pmatrix} \ddot{\theta}_0(t) \\ \ddot{\theta}_1(t) \end{pmatrix} + \begin{pmatrix} \delta_0 + \frac{1}{4}m_1l_1^2\dot{\theta}_1(t)\sin 2\theta_1(t) & -\frac{1}{2}m_1l_0l_1\dot{\theta}_1(t)\sin\theta_1(t) \\ -\frac{1}{8}m_1l_1^2\dot{\theta}_0(t)\sin 2\theta_1(t) & \delta_1 \end{pmatrix} \begin{pmatrix} \dot{\theta}_0(t) \\ \dot{\theta}_1(t) \end{pmatrix} + \begin{pmatrix} 0 \\ -\frac{1}{2}m_1g l_1 \sin\theta_1(t) \end{pmatrix} = \begin{pmatrix} M(t) \\ 0 \end{pmatrix} \quad (4)$$

where m_0 , m_1 stand for the masses of the arm and the pendulum, l_0 , l_1 are their respective lengths, δ_0 , δ_1 are the damping constants in the joints of the arm and pendulum, $J_0 = \frac{1}{3}m_0l_0^2$ and $J_1 = \frac{1}{3}m_1l_1^2$ are the moments of inertia of the arm and pendulum with respect to their pivot points and $M(t)$ is the input torque applied upon the rotary arm.

Both models listed above (hereafter referred to as *force/torque models*) are among those which were included in the *Inverted Pendula Models* sublibrary of the *IPMaC* [2] in form of atomic library blocks: the *Classical Single Inverted Pendulum* and the *Rotary Single Inverted Pendulum* block.

B. Modeling and Simulation of Actuating Mechanisms

To provide the simulation models of IPS with a model of the most frequently used actuating mechanism, a library block *DC Motor for Inverted Pendula Systems* was included into the the *Inverted Pendula Motors* sublibrary of the *IPMaC*. The block implements the mathematical model of a brushed direct-current (DC) motor in two alternative forms depending on the type of system (classical / rotary) [7]:

- a voltage-to-force conversion relationship:

$$F(t) = \frac{k_m k_g}{R_a r} V_a(t) - \frac{k_m^2 k_g^2}{R_a r^2} \dot{\theta}_0(t) \quad (5)$$

- and a voltage-to-torque conversion relationship:

$$M(t) = \frac{k_m k_g}{R_a} V_a(t) - \frac{k_m^2 k_g^2}{R_a} \dot{\theta}_0(t) \quad (6)$$

where $V_a(t)$ is the input voltage applied to the motor, k_m is the motor torque constant, equal in value to the back EMF constant, k_g is the gear ratio, R_a is the armature resistance and r is the radius of the tooth pulley which is coupled to the motor shaft and converts the torque produced by the motor into the linear driving force $F(t)$ (hence, r is only necessary for classical IPS). If the DC motor model is appended to an inverted pendulum system (i.e. (5) is substituted into (3), or (6) is substituted into (4)), a *voltage model* of the system is obtained.

III. OUTLINE OF A CONTROL STRATEGY FOR SWING-UP AND STABILIZATION OF SINGLE INVERTED PENDULUM SYSTEMS

As a principal control objective, the pendulum had to be *swung up from the stable downward to the unstable upright equilibrium* ($\mathbf{x}(t) = \mathbf{0}^T$), *captured and stabilized there* [8].

This problem leads to a hybrid control setup which consists of three basic components, schematically depicted in the block diagram in Fig. 3:

- a *swing-up controller*, which actuates the base with an input signal of appropriate direction and magnitude, making the pendulum swing with an increasing amplitude and angular speed until it enters the stabilization zone (balancing region) around the upright position [1],
- a *stabilizing (balancing) controller*, which maintains the pendulum in the upright position using suitable linear or nonlinear feedback control techniques [2][4][5][9],
- a *transition (switching) mechanism*, which intercepts the pendulum when it nears the upright position (crosses the borderline of the balancing region) and switches to stabilizing control [8].

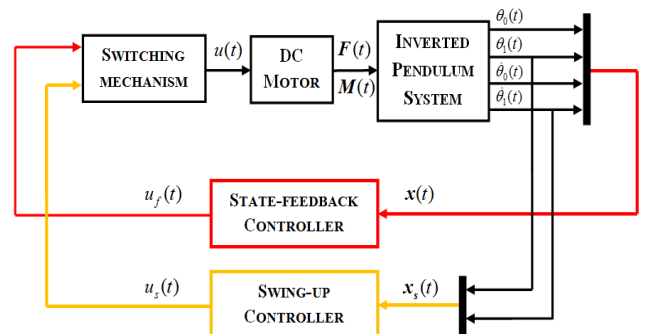


Fig. 3 Block diagram of swing-up and stabilization control of IPS

A. Energy-Based Methods for Swing-up Control

One of the earliest and most effective approaches to pendulum swing-up is the method based on energy considerations, proposed and explained by Åström, Furuta and Iwase in [8][10]. The goal is to maximize the total mechanical energy W of the pendulum in the upright position. The first derivative of the energy is given as

$$\dot{W} = \frac{d}{dt} \left(\frac{1}{2} m l^2 \dot{\theta}_1^2(t) + m g l (1 - \cos \theta_1(t)) \right) = m l \dot{\theta}_1(t) \cos \theta_1(t). \quad (7)$$

Whether is greater or lower than zero depends directly on the sign of $\dot{\theta}_1(t) \cos \theta_1(t)$. The basic energy-based control law (cosine value controller) therefore becomes [11]:

$$u(t) = -u_m \operatorname{sgn}(\dot{\theta}_1(t) \cos \theta_1(t)) \quad (8)$$

This control law can be simplified into a law which will be referred to as a *zero speed controller*:

$$u(t) = u_m \operatorname{sgn}(\dot{\theta}_1(t)) \quad (9)$$

or modified into an *absolute value control law*:

$$u(t) = u_m \operatorname{sgn}(\dot{\theta}_1(t) |\theta_1(t)|) \quad (10)$$

The control laws (8)-(10) were encapsulated into the structure of *Swing-up Controller* block of the *IPMaC*, which allows the user to select a swing-up method, the constraints of the balancing region and the input magnitude u_m .

B. Stabilization via State-Feedback Control Techniques

Once the pendulum has been captured in the upright position, we need to switch to a balancing controller. The *Inverted Pendula Control* sublibrary of the *IPMaC* provides complex software support for the *linear state-feedback (S-F) controller design*. The blocks it contains (*S-F Controller with Feedforward Gain*, *S-F Controller with Integral Action* and *Luenberger Estimator*, among others) were thoroughly described in [2][4] in terms of their structure and functionality.

With aid of the mentioned blocks, the feedback gain vector which brings a linear system into the state space origin can be determined alternatively through pole-placement and linear quadratic regulator (LQR), using either the continuous-time or discrete-time linearized state-space models of IPS. These can be obtained from the *Inverted Pendula Model Linearizator & Discretizer* (Fig. 4), a MATLAB GUI application which generates the state-space matrices by expanding (2) into the Taylor series. Additional control structures ensure that the output reaches a reference position by means of feedforward gain and any permanent steady-state error is eliminated by the implemented integral control.

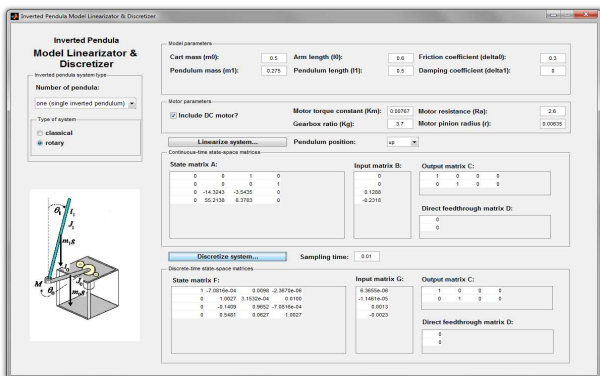


Fig. 4 *Inverted Pendula Model Linearizator & Discretizer*

IV. EVALUATION OF THE CONTROL STRATEGY

The proposed control strategy was verified for the *force/torque models* and *voltage models* of both classical single and rotary single IPS. The numeric parameters were chosen to be $m_0 = 0.5 \text{ kg}$, $m_1 = 0.275 \text{ kg}$, $l_0 = 0.6 \text{ m}$, $l_1 = 0.5 \text{ m}$, $\delta_0 = 0.3 \text{ kg s}^{-1}$, $\delta_1 = 0 \text{ kg m}^2 \text{ s}^{-1}$, and the initial conditions were set to $\theta_0(0) = -1$ and $\theta_1(0) = -\pi$. The DC motor model parameters were borrowed from the motor featured within the series of popular laboratory models of IPS issued by *Quanser Academic* [7]. The weight matrices of the standard discrete-time LQ functional $J_{LQR}(i) = \sum_{i=0}^{N-1} x^T(i) Q x(i) + u^T(i) r u(i)$ for the balancing state-feedback controller were specified as $Q = \operatorname{diag}(100 \ 20 \ 20 \ 0)$, $r = 1$ for every simulation.

Fig. 5 and Fig. 6 depict the comparison of the three swing-up control laws (8)-(10) applied on the force/torque models of IPS. For each system and swing-up control law, the input magnitude was tuned to the highest value which still allows successful swing-up with no oscillations or destabilization (Tab. 1). Switching between the controllers took place when the pendulum was 0.6 rad away from the upright position.

TABLE I HIGHEST INPUT MAGNITUDES WHICH ENSURE SUCCESSFUL SWING-UP		
Type of swing-up controller	classical single inverted pendulum (force [N])	rotary single inverted pendulum (torque [Nm])
absolute value controller	2,1	2,2
zero speed controller	2,6	2
cosine value controller	1,6	1,3

In the case of *classical inverted pendulum*, the best performance in terms of both pendulum swing-up time and

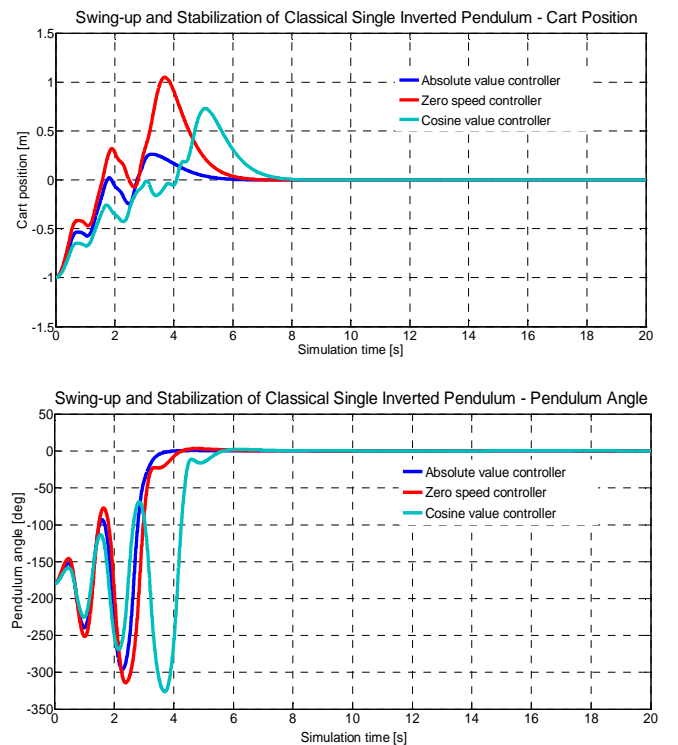


Fig. 5 Swing-up and stabilization of the classical single inverted pendulum force model – comparison of methods

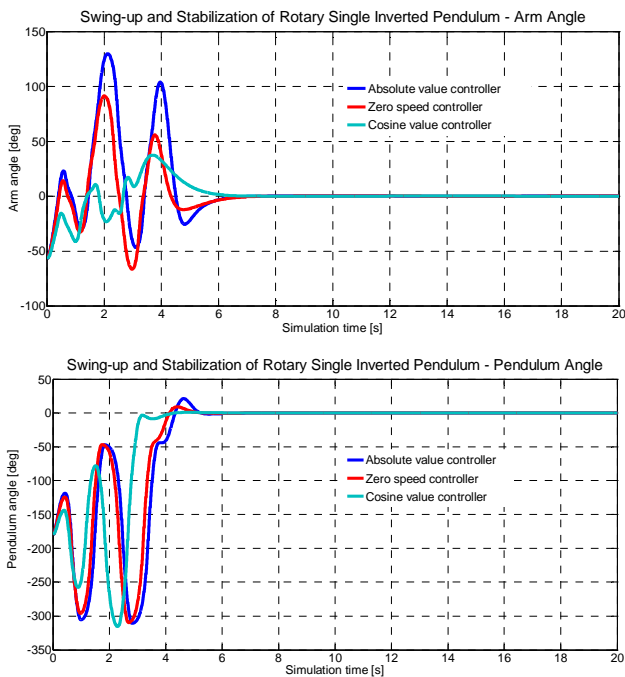


Fig. 6 Swing-up and stabilization of the rotary single inverted pendulum torque model – comparison of methods

total displacement of the cart was provided by the absolute-value controller. The zero speed controller only managed to stabilize the pendulum at the comparable time at the price of a large cart overshoot. The *rotary inverted pendulum* performed at its best for the cosine-value controller – the performance was diminished for other controllers due to long intervals during which energy is taken from the pendulum.

When tuning the input magnitude u_m , the time needed to swing the pendulum to the upright position was discovered as indirectly proportional to u_m . However, feasible input values could only be selected from a bounded interval. An unlimited increase in magnitude leads to the increase of the pendulum

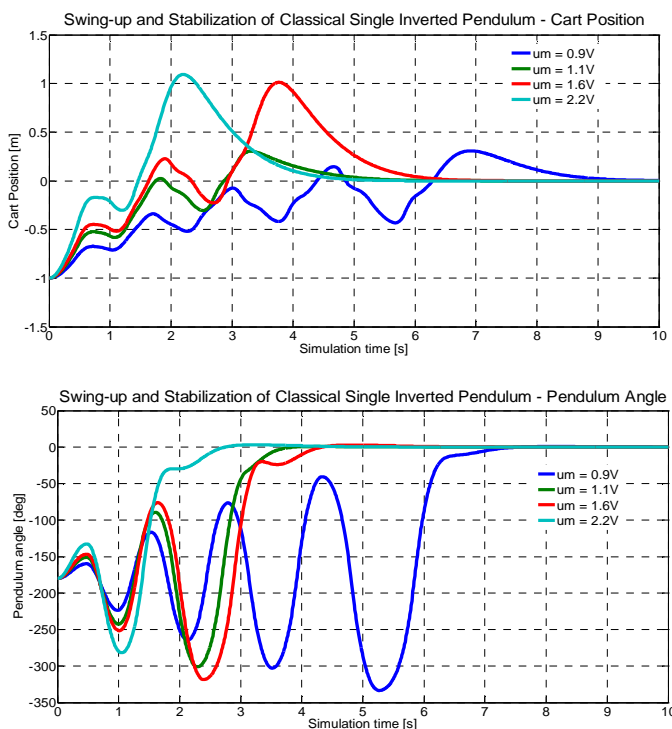


Fig. 7 Swing-up and stabilization of the classical single inverted pendulum voltage model – comparison of the effect of varying input magnitude

angular speed, and a too low magnitude may either be entirely unable to bring the pendulum upright, or it needs numerous swings to do so, which results in a displacement of the base away from the equilibrium point. In both cases, the control mechanism would fail to stabilize the system, proving that the balancing controller is only effective if the state of system is close to the equilibrium point in the moment of interception. An illustrative experiment which compares the effect of a varying u_m on the classical single inverted pendulum voltage model, is depicted in Fig. 7.

V. CONCLUSION

The purpose of this paper was to provide a thorough overview of the problem of swing-up and stabilizing control design for one-link inverted pendula systems. Several swing-up methods based on energy considerations were proposed and compared in simulation experiments involving models of a classical and rotary single inverted pendulum; linear state-feedback control techniques were next used to stabilize the pendulum, captured in the upright position. Most importantly, it was shown that the *Inverted Pendula Modeling and Control*, the custom-designed *Simulink* block library which was used as a software framework for all issues covered in the paper, is a highly suitable program tool for design and verification of hybrid control methods for nonlinear systems.

ACKNOWLEDGMENT

This work has been supported by the Scientific Grant Agency of Slovak Republic under project Vega No.1/0286/11 Dynamic Hybrid Architectures of the Multiagent Network Control Systems.

REFERENCES

- [1] K. Furuta, M. Yamakita, S. Kobayashi: "Swing Up Control of Inverted Pendulum", *Proc. of the Int. Conf. on Industrial Electronics, Control and Instrumentation (IECON'91)*, Oct 28-Nov 1, 1991, Kobe, Japan
- [2] S. Jadlovska, J. Sarnovský, "An extended Simulink library for modeling and control of inverted pendula systems." *Proc. of the Int. Conf. Technical Computing Prague 2011*, November 8, 2011, Prague, Czech Republic, ISBN 978-80-7080-794-1
- [3] S. Jadlovska, A. Jadlovska, "Inverted pendula simulation and modeling – a generalized approach," *Proc. of the 9th Int. Scientific-Technical Conf. on Process Control*, June 7-10, 2010, University of Pardubice, Czech Republic, ISBN 978-80-7399-951-3
- [4] S. Jadlovska, J. Sarnovský, "Classical double inverted pendulum – a complex overview of a system," *Proc. of the IEEE 10th Int. Symposium on Applied Machine Intelligence and Informatics – SAMI 2012*, January 26-28, 2012, Herľany, Slovakia, ISBN 978-1-4577-0195-5
- [5] M. Schlegel, J. Mešťánek, "Limitations on the inverted pendula stabilizability according to sensor placement," *Proc. of the 16th Int. Conf. on Process Control*, June 11-14, 2007, Štrbské Pleso, Slovakia, ISBN 978-80-227-2677-1
- [6] P. Ernest, P. Horáček, "Algorithms for control of a rotating pendulum", *Proc. of the 19th IEEE Mediterranean Conf. on Control and Automation (MED'11)*, Corfu, Greece, 2011
- [7] Quanser Academic, Rotary Motion Servo Plant: SRV02, User Manual, No. 700, Rev. 2.3
- [8] K.J. Åström, K. Furuta: "Swinging up a pendulum by energy control," *Proc. of the 13th IFAC World Congress*, June 30 - July 5, 1996, San Francisco, California 1996
- [9] Kats, C.J.A. (2004): Nonlinear control of a Furuta rotary inverted pendulum, DCT Report, No. 2004:69
- [10] K. Furuta, M. Iwase, "Swing-up time analysis of pendulum," in *Bulletin of the Polish Ac. of Sciences – Technical Sciences*, vol. 52, no. 3, 2004
- [11] M. Bugeja, "Non-linear swing-up and stabilizing control of an inverted pendulum system," *Proceedings of the IEEE Int. Conf. on Computer as a Tool (EUROCON'03)*, Sep. 2003, Ljubljana, Slovenia

Tool Support for Task-driven Case Study Design and Evolution

¹Milan NOSÁL (*1st year*)

Supervisor: ²Jaroslav PORUBÄN

^{1,2}Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹milan.nosal@tuke.sk

Abstract—This paper presents tool support for scenarios' update in task-driven case studies. Firstly there is a brief introduction to task-driven case study methodology and scenarios are presented. Motivation for automating scenarios' update is stated. The paper elaborates the proposed and implemented solution on examples from Minesweeper task-driven case study. Solution is based on attribute-oriented programming and generative programming.

Keywords—case study, task-driven methodology, attribute-oriented programming, generative programming

I. INTRODUCTION

Computer science and other computer related subjects affects virtually all areas in life. Computers play significant role in industry, business and nowadays even personal life of individual. This interest in information technologies (IT) also affects the labour market, offering many IT related positions. We want to focus especially on software engineering (SE) and programming courses. Decade ago, Stiller et al. [9] was pointing out that large-scale software development fiascoes were still quite prevalent, despite years of requiring at least one software engineering course in accredited computer science programs. Also Goldfinch [4] five years later bears witness of many failures in software development. This experiences show that there is still need for more quality in SE courses.

In many of our programming and SE courses we use task-driven case studies on seminars. This approach is proving itself valuable in education process. In this approach educators use task-driven scenarios in a form of xHTML web pages, one for each seminar in the course, that guide students through case study. In the programming and software engineering courses the case studies are usually software projects to be developed. In this paper we present tools we developed to support process of designing a case study and corresponding scenarios.

II. BACKGROUND

This section gives introduction to task-driven case study approach. As its name suggests, the approach is based on two concepts, task-driven methodology and case studies.

A. Task-driven teaching

To see once is better than to hear hundred times.
– Slovak Proverb

This proverb speaks about engaging student in learning process. Better than hear how to do something is to see how it is done. And even better is to try it personally. Traditional teaching model, "transmission-accept" model considers

that the knowledge and technology are acquired through the teacher's teaching [2] - hearing the theory. Giving students theoretical knowledge base in lectures is surely necessary, but it is presenting itself as insufficient to provide understanding that is needed in order to apply knowledge in practice.

We believe that learning on seminars should consist of tasks for students to involve them in learning process. We are not alone in this belief, many educators use task-driven methodology that is based on tasks [2], [10], [8], [11]. Students have tasks they need to solve by themselves and therefore can not wait for teacher to do it instead of them.

As Peng et al. [8] says, this way we can change students' view of learning. Instead of seeing learning as a task, they start learning because of the tasks. This changes their attitude from passively accepting knowledge to pursuing it for the needs of the tasks. This learning model mimics better real world practice. In their future jobs, they will be given tasks and it will be their responsibility to get required knowledge for solving them.

B. Case studies

The issue with so far most common academic projects is that projects may not present a highly abstracted view of reality and students usually focus on the project outcome (product) and not the process. The process of implementing project is the part of the course in which they should learn how to apply theoretical knowledge. Thus competence of a student in applying skills or synthesis of learning in real world scenarios should not be established from a typical academic project based assignment, but rather from a case study designed by experienced professional. [3]

A case study is a method of involving students in real life scenarios. It investigates phenomena in real-life context [5]. An engineering case is an account of an engineering activity or problem containing some of the background and complexities usually encountered by an engineer. It provides the proper context of the problem to be solved and problem's description. It should also contain information regarding the processes under consideration, the technologies and tools used. [1]

Anwar [1] states main reasons for using case studies. We find the most important the first reason – introducing more complex problems into the classroom, ones that are closer to real world instead of typical introductory examples. I remember my course of object-oriented programming. We were taught about inheritance and on practical seminars we were trying it out on simple example of Ship class, and Cruiser

and Yacht classes inheriting from Ship. It showed me how I can implement it, but there was no context for its usage. It is easy to teach how to implement something, but much harder is to give the problem sufficient context, so the students will understand when and why they can use the taught principle [6]. We believe that we can bring this context into classroom by using case studies.

III. TASK-DRIVEN CASE STUDY IN JAVA TECHNOLOGIES COURSE

Combining task-driven methodology with case studies we get task-driven case studies. In this approach, students are given a case study, that is developed by solving tasks. These tasks help students to get good practices and get into touch with every area of knowledge that is presented in the course. Classical case study gives the student context to course, but can not ensure that student will solve it right way. Of course there has to be space for mistakes and learning from them, but educators can not let students do it their own way. Imagine the project of Object-oriented programming course implemented procedurally (whole in one or two classes using static methods) – just because the student was familiar with procedural programming and wasn't willing to take harder path of learning. Of course, educator can fail him, but would it be better than just making him take the right path?

Task-driven case study was first used in Java Technology course. The case is a simple game – Minesweeper. Decision of using a game as case study is based on philosophy of learning by playing. Using the words of Mingins et al. [7]: “the course should be fun! To encourage the students to continue their studies in this area the material needs to be interesting and stimulating and needs to maintain a balance between the whole cohort of students implementing the same specification and at the same time each having the freedom to develop new and interesting extensions to its project.” Using a game has a clear goal – motivate students to get to work.

Although a case is a simple game, it uses graphical user interface (GUI). For programmatically unexperienced students this poses a challenge but with sufficient award for solution – they get their own implementation of the game with GUI. Students get basic framework of the Minesweeper source codes and case study scenarios that describe tasks they have to solve in context of the Minesweeper case. These scenarios are supposed to guide them through application of all taught Java technologies (or other principles in general according to course syllabus). Following sections will be backed up by examples from this task-driven case study.

A. Case study scenarios

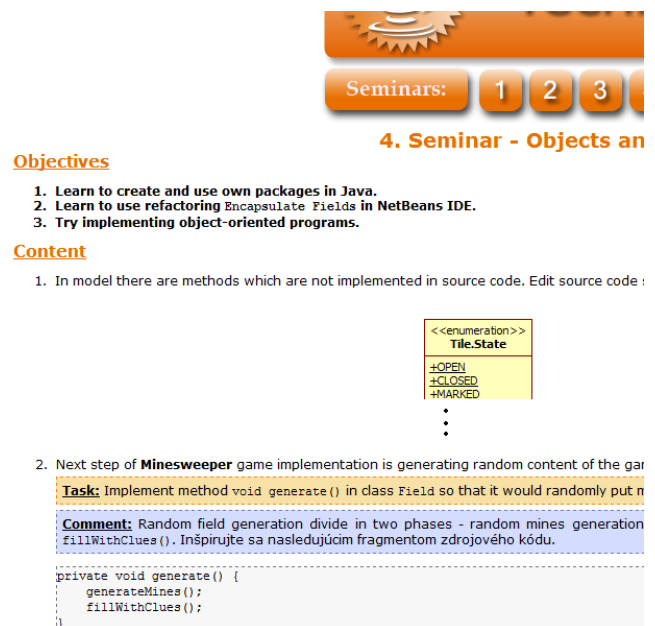
For scenarios definition we use XML language. Its main elements are `<module>`¹ and `<task>`. Whole case study is divided into modules representing main tasks. Each module corresponds with one seminar and it should be consistent with theory presented on last lecture. Module should have clear education objectives. All the objectives have to be covered by tasks in given module. XML language is used simply to separate content from presentation form.

¹From now on we will use brackets `<>` to indicate we are speaking about XML element and `@` sign to indicate speaking about XML attribute. Annotations in Java start also with `@` sign, but we will omit it when referring to annotations to avoid misunderstanding.

In fig. 1 there is shown basic structure of scenario introducing objects and classes in Java. Whole module is divided in few steps that may contain tasks for students to solve. Each step aims to fulfil one or more of the objectives. Step may contain text, images or videos introducing some aspect of the case study, or some theoretical knowledge needed for next step – it does not have to contain any task if it is not necessary. Tasks indicate students that they have to work on solving them. Other content is context for these tasks. Fig. 2 shows example of presented scenario as a web page.

```
<module id="02">
  <title>Objekty a triedy</title>
  <objectives>
    <objective>..</objective>
    ..
  </objectives>
  <content>
    <step>
      ..
      <task id="getValue">
        ..
      </task>
      ..
    </step>
  </content>
</module>
```

Fig. 1. Basic structure of XML scenarios showing how the content is created.



4. Seminar - Objects and Classes

Objectives

1. Learn to create and use own packages in Java.
2. Learn to use refactoring Encapsulate Fields in NetBeans IDE.
3. Try implementing object-oriented programs.

Content

1. In model there are methods which are not implemented in source code. Edit source code :


```
<<enumeration>>
Tile.State
+OPEN
+CLOSED
+MARKED
...
```
2. Next step of Minesweeper game implementation is generating random content of the game field.

Task: Implement method void generate() in class Field so that it would randomly put n mines in the field.

Comment: Random field generation divide in two phases - random mines generation and random numbers generation. Inšpirujte sa nasledujúcim fragmentom zdrojového kódu.

```
private void generate() {
    generateMines();
    fillWithClues();
}
```

Fig. 2. Screenshot of one scenario of the Minesweeper case study that shows actual presentation form.

IV. MOTIVATION

Of course the case study project needs to be implemented first to verify its suitability for elaboration by students. Thus scenarios design is driven by the case study implementation. Author has to divide the case study to tasks and to give them proper formulation. These tasks are divided into scenarios. They can be written after implementing a case study, or simultaneously with it. In both cases author has to work with two different sources of information about a case study, he has to jump from source codes to XML and vice versa, because there is direct connection between tasks and source code. Tasks' solutions are parts of the case study's source code.

Design process of task-driven case study is challenging and time consuming process. This necessity of using two different views (or even editors) on source code and on scenarios makes it even slower. Fig. 3 shows getter for an instance variable in Minesweeper project. For this getter, scenarios' author has to define a task as in fig. 4. While the getter is in the source code, the task is in the XML document for corresponding module.

```
public int getRowCount() {
    return rowCount;
}
```

Fig. 3. Getter for rowCount in Minesweeper implementation.

```
<task id="getter">
    Implement getter for instance variable rowCount.
</task>
```

Fig. 4. Task for getter shown in fig. 3.

There is very close proximity between source code and scenarios. Keeping this two elements consistent is an important challenge not only in design process, but also during evolution of the case study in time of its usage.

V. TOOL SUPPORT FOR TASK-DRIVEN CASE STUDY DESIGN AND EVOLUTION

We believe that the design process and case study evolution would be easier and faster if source code and scenarios would be loosely coupled. Task definitions that concern source code directly should be defined in source code and injected into scenarios automatically. To realize this vision we decided to use custom Java annotations (implementation of attribute-oriented programming on the Java platform).

Fig. 5 sketches the overview of update process elaborated in this section. Loop between XSLT template for updating scenarios and XML scenarios express real connection between them. XSLT template never creates new scenario, it merely updates existing. Depending on the situation it updates empty templates (design of new case study) or old scenarios (evolution of existing case study).

A. Source code task annotations

We designed three annotations so far:

- Task - annotates source code elements (as class or method). It has properties that define the module and its task to which the source code element belongs. Both module and task has an identifier.
- TaskDescription - allows to add content to task body through the source code.
- Comment - allows to add a comment to the task.

These annotations are used to annotate source code of case study. They define the division of source code into tasks and modules (there are so many modules, as many different module identifiers are used in Task annotations). Task definition in source code corresponding with task outlined in fig. 4 is shown in fig. 6.

```
@Task(module = "02", id = "getter")
@TaskDescription(value="Implement
getter for instance variable rowCount.")
public int getRowCount() {
    return rowCount;
}
```

Fig. 6. Annotated getter for rowCount in Minesweeper implementation.

B. Annotation processing

Annotated source code is processed by annotation processor TaskProcessor. This processor creates XML representation of tasks, their descriptions and corresponding source code elements' declarations. In fig. 7 there is an XML representation of getter task generated from code in fig. 6.

```
<module id="02">
  <task id="getter">
    <method>
      <name>getRowCount</name>
      <full_name>getRowCount ()</full_name>
      <public/>
      <enclosing_type>
        minesweeper.core.Field
      </enclosing_type>
      <return_type>int</return_type>
      <desc>Implement getter for instance
        variable rowCount.</desc>
    </method>
  </task>
  ..
```

Fig. 7. XML representation of getter task.

Although the XML scenarios can be modified directly, we chose to use this inbetween step to abstract the approach from programming language. If the case study is not implemented in Java, we merely need to implement new annotation processor for given platform. As fig. 6 and fig. 7 show, the mapping between annotated source code and its XML representation is pretty straightforward and therefore implementing such processor is not a hard task.

C. Updating scenarios according the source code

Updating an empty scenario template (without tasks) in case study design process or existing scenario during the evolution is realized using XSLT processor. We designed XSLT template that takes XML representation of source code and merges it with scenario template (it does not matter whether the template is empty or already filled with tasks).

As the XSLT transformation has to deal with evolution too, therefore we needed to solve the problem of finding the right XML elements to update. Of course, one task may span through more source code elements (one task may include implementing more methods), therefore there is used identifier to determine the source of XML element. In fig. 8 there is a task with identifier "markTile". All Task annotations with "markTile" identifier belong to this element.

Descriptions provided in TaskDescription annotations are mapped to elements <generated> that are children of relevant <task> element. TaskDescription can annotate only source code element that is annotated with Task. It is mapped to <generated> element with an attribute @from identifying source code element that is

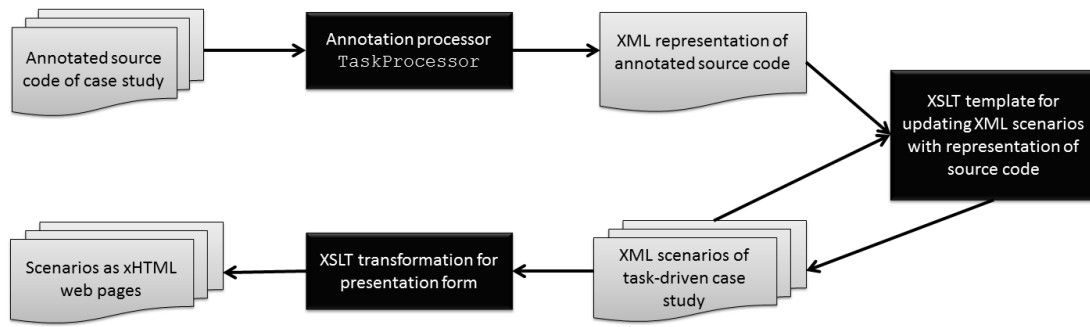


Fig. 5. Scheme of data flows in the update process of scenarios from source code using tool support.

```

<task generated="true" id="markTile">
  <generated
    from="minesweeper.core.
      Field.markTile(int,int)">
    Implement markTile.
  </generated>
  ..
</task>

```

Fig. 8. Task for getter shown in fig. 3 generated by the tool.

annotated with corresponding TaskDescription². Pseudocode for algorithm of generating <task> from XML representation of sources is outlined in fig. 9. For example, when there is TaskDescription annotation on method markTile(int,int) belonging to "markTile" task, we expect that the output of generation would result in XML as in fig. 8.

```

generate <task @generated="true">
  with corresponding @id;
foreach <desc> in <task> with corresponding @id
  in XML repr. of sources {
    generate <generated>
      with corresponding @from to <task>;
    generate content for <generated> from <desc>;
  }
foreach <comment> in <task> with corresponding @id
  in XML repr. of sources {
    generate <comment @generated="true">
      with corresponding @from to <task>;
    generate <generated>
      with corresponding @from to <comment>;
    generate content for <generated> from <comment>;
  }

```

Fig. 9. Pseudocode for basic <task> generation.

In process of generation, all <task> elements are scanned and if there is a Task annotation³ that does not have a corresponding XML element, the new <task> is generated (algorithm in fig. 9). Old <task> elements are updated according to annotations and when some annotation was removed, the corresponding element is removed too. In fig. 8 one may notice the @generated attribute set to "true" in the <task> element. This flag is used to indicate whether the task should be regenerated in next XSLT transformation. If it is present and set to true, XSLT transformation is allowed to process

its content and update it. In other cases the transformation leaves it untouched (even if it does not have corresponding annotation). This mechanism gives the user freedom to update only parts of the scenarios that he wants to update. Whole process is outlined in fig. 10.

```

if there is Task in sources
  for which there is not <task>
    generate new <task> from Task
if there is <task @generated="true">
  for which there is not Task
  delete this <task>
if there is <task @generated="true">
  with corresponding Task
  update <task> according to Task
if there is <task>
  without @generated="true"
  with corresponding Task
  leave <task> untouched

```

Fig. 10. Updating <task>s.

In the similar manner works generating of comments. Using generated elements for description allows the user of the tool to add custom content in task element. If this content is not put into generated element, it remains untouched even when generated flag of the task is set to true. This allows user to extend scenario with the content that is not close to source code and should not be present at source code annotations.

VI. CONCLUSION

Presented tool support allows task-driven case study author to put source code dependant task information directly in the source code. The scenarios are automatically updated later using chain of tools as presented in fig. 5. This idea allows easier and safer evolution of task-driven case study, because using the tool we can not forget to update some task according to the case study. Task annotations in source code allows also better understanding of the case study, because all important information about case study are centered in source code and not scattered between source code and XML documents.

ACKNOWLEDGEMENT

This work is the result of the project implementation: IT4KT - Information Technologies for Knowledge Transfer (ITMS project code: 26220220123) supported by the Research & Development Operational Program funded by the ERDF.

²@from in XML scenarios link XML elements to annotated source code elements.

³Of course, this information is retrieved from XML representation of source code, not directly from sources.

REFERENCES

- [1] Anwar S., "Use of engineering case studies to teach associate degree electrical engineering technology students", in Proceedings of the 31st Annual Frontiers in Education Conference, Vol.3, pp. 8–10, 2001.
- [2] Dong Y., "A Graded Task-driven Methodology for Computer Science Education", in Second International Workshop on Education Technology and Computer Science (ETCS), Vol. 3, pp. 654–656, 6–7 March 2010.
- [3] Garg K., Varma V., "Case Studies as Assessment Tools in Software Engineering Classrooms", in Proceedings of the 22nd Conference on Software Engineering Education and Training, pp. 8–11, 17–20 February 2009.
- [4] Goldfinch S., "Pessimism, Computer Failure, and Information Systems Development in the Public Sector", in Public Administration Review, Vol. 67, Issue 5, pp. 917–929, 2007.
- [5] Jia Y., "Improving software engineering courses with case study approach", in Proceedings of the 5th International Conference on Computer Science and Education (ICCSE), 2010, pp. 1633–1636, 24–27 August 2010.
- [6] Martin F., "Toy projects considered harmful", in Commun. ACM, Vol. 49, Issue 7, pp. 113–116, July 2006.
- [7] Mingins C., Miller J., Dick M., Postema M., "How We Teach Software Engineering", in JOOP, Vol. 11, No. 9, pp. 64–66, 74, 1999.
- [8] Peng W., Jingjing X., "The implementation and harvests of task-driven in basic computer education at university", in Proceedings of International Conference on E-Health Networking, Digital Ecosystems and Technologies (EDT), vol.2, pp. 311–314, 17–18 April 2010.
- [9] Stiller E., LeBlanc C., "Effective software engineering pedagogy", in J. Comput. Small Coll., Vol. 17, Issue 6, pp. 124–134, May 2002.
- [10] Xue J., Zhang L., "Application of task-driven approach in information technology education", in Proceedings of International Conference on Electrical and Control Engineering (ICECE), pp. 2024–2026, 16–18 September 2011.
- [11] Yu D., Wang Q., "Task-Driven Method in Practical Teaching of Software Engineering", in Proceedings of the Third Pacific-Asia Conference on Circuits, Communications and System (PACCS), 2011, pp. 1–3, 17–18 July 2011.

Section:
Informatics & Telecommunications
Poster form

Acoustic Event Detection using MFCC and Genetic Algorithm

Eva VOZÁRIKOVÁ (3rd year)

Supervisor: Anton ČIŽMÁR

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

eva.vozarikova@tuke.sk

Abstract—This paper presents the detection of abnormal situation via sound information. It is focused on the detection of breaking glass and gun shots. As a main feature extraction algorithm MFCC, statistical moments and ZCR were used. For improving the recognition score the feature selection algorithm called Minimum Redundancy Maximum Relevance was applied. The proposed framework of recognizing potentially dangerous acoustic events through Hidden Markov Models based classification is presented here.

Keywords—Acoustic events, MFCC, feature selection, MRMR

I. INTRODUCTION

In last few years a huge interest was concentrated to the detection of selected sound, where one category of sounds can include potential dangerous ones, for example gun shots, sounds of car crash accidents, calling for help, etc. The detection of the mentioned sound category is relevant for surveillance or security systems. They are created these times by using the knowledge of different scientific areas, e.g. signal processing, pattern recognition, machine learning, etc. As it was indicated, surveillance or security systems can help to protect lives and properties. In some cases these systems can be very helpful for police because they operate according to the given instructions and they are not affected by any irrational behavior in comparison with people, e.g. the bystander effect [1], [2].

Each surveillance system can be divided into the feature extraction block and the classification block. In the feature extraction block, the input audio signal is transformed into the feature vectors according to the chosen feature extraction algorithm in time, frequency or cepstral domain.

In this paper the feature extraction based on Mel-Frequency Cepstral Coefficients - MFCC [3], which were primary developed for speech or speaker recognition tasks together with statistical moments (skewness and kurtosis) and Zero Crossing Rate - ZCR were performed. Interesting results were achieved also by the basic spectral descriptors defined in MPEG-7 standard [4], [5].

The efficient feature extraction should highlight the relevant information and reduce the number of input data by removing irrelevant information using various methods. Discrete Cosine Transform - DCT is used in MFCC extraction for decorrelating coefficients. Other kinds of methods such as Principal Component Analysis - PCA, Linear Discriminant Analysis - LDA, etc. transform original feature space to another space

with lower dimension. A feature selection algorithm such as Minimum Redundancy Maximum Relevance - MRMR, Forward Selection - FS, Backward Selection - BS, ReliefF, etc. can be used to select relevant subset of features from full set of features. Some of them (PCA, LDA) change the values of coefficients, other (FS, BS, ReliefF) select the relevant ones or sort them to the new sequence according to their priority (MRMR).

In the block of classifier the input patterns are classified into predefined classes according to the knowledge obtained in the training process of classifier. There are several methods that can perform it e.g. Hidden Markov Models (HMM), Gaussian Mixture Models (GMM), Bayesian Networks (BN), Support Vector Machines (SVM) and Neural Networks (NN), etc. In this work one, two, three and four states HMMs up to 1024 PDFs (Probability Density Functions) were used.

The research described in this paper is focused on the feature extraction in the time and cepstral domain. The cepstral domain is represented with Mel-Frequency Cepstral coefficients - MFCC, and the time domain is represented with the statistical moments (skewness, kurtosis) and Zero Crossing Rate - ZCR. In the post processing phase of MFCC the selection of relevant subsets of feature were done by MRMR - Maximum Relevance Minimum Redundancy algorithm.

We have supposed that information in MFCC can be reduced without a decrease system performance. For sorting the coefficients MRMR algorithm was used [7], which is mostly used in genetic. We applied the same approach, which was used for selection of relevant genes, for selection the relevant subset from full set of MFCC features. This approach is applied to the acoustic event detection framework as a partial task in the complex surveillance system.

The rest of paper has following structure: section 2. presents feature extraction methodology. Section 3. gives information about used feature selection algorithm - MRMR and section 4. describes the selection process and performed experiments. Section 5. summarizes the obtained results.

II. FEATURE EXTRACTION

The effective feature extraction is a very important phase of overall process, because the recognition performance directly depends on the quality of extracted feature vectors. In this work, extraction of MFCC, skewness, kurtosis and ZCR were done. The theoretical background is described below.

A. Mel-Frequency Cepstral coefficients - MFCC

Ear's perception of the frequency components in the audio signal does not follow the linear scale, but rather the Mel-frequency scale, which should be understood as a linear frequency spacing below 1 kHz and logarithmic spacing above 1 kHz. So filters spaced linearly at a low frequency and logarithmic at high frequencies can be used to capture the phonetically important characteristics. The relation between the Mel-frequency and the frequency is given by the formula:

$$Mel(f) = 2595 \times \log_{10} \left(1 + \frac{f}{700} \right), \quad (1)$$

where f is frequency in Hertz. MFCC coefficients are computed following way: a segment of signal is divided into the 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 Discrete Fast Fourier Transform - DFFT, 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 [6].

B. Skewness and Kurtosis

Skewness and kurtosis are statistical measures that describe the shape of data set.

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. It is described by the formula:

$$Skewness = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^3}{(N-1)s^3}, \quad (2)$$

where Y_1, Y_2, \dots, Y_N are sample of data set, \bar{Y} is the mean, s is the standard deviation, and N is the number of data points. The skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero.

Kurtosis is a measure of the "peakedness" of the probability distribution and it is described by the formula:

$$Kurtosis = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^4}{(N-1)s^4}, \quad (3)$$

where Y_1, Y_2, \dots, Y_N are sample of data set, \bar{Y} is the mean, s is the standard deviation, and N is the number of data points. The kurtosis for a standard normal distribution is equal to three.

C. Zero Crossing Rate

Zero Crossing Rate - ZCR is a very useful audio feature, it is defined as the number of times that the audio waveform crosses the zero axis. It is defined by the formula:

$$ZCR = \frac{1}{2} \left(\sum_{n=1}^{N-1} [sign(s(n)) - sign(s(n-1))] \right) \frac{Fs}{N}, \quad (4)$$

where N is total samples of the signal $s(n)$, F_s is a sampling frequency and $sign(x)$ function is defined follows:

$$sign(x) = \begin{cases} 1 & \text{ak } x > 0, \\ 0 & \text{ak } x = 0, \\ -1 & \text{ak } x < 0. \end{cases}$$

III. FEATURE SELECTION

A feature extraction can consist from two different steps, more precisely extraction of characteristic features according to the extraction algorithm, then the selection of relevant subset of features. In some works, two phase feature selection was applied [7], where first algorithm selects candidate feature set and next one selects final subset of features. Some of algorithms need to cooperate with particular classifier, others work for overall classifier. Minimum Redundancy Maximum Relevance - MRMR belongs in to the last group.

MRMR algorithm [7],[8] selects a compact set of superior features at very low cost. Algorithm includes two independent criterions:

- *maximal relevance*,
- *minimal redundancy*.

First criteria *maximal relevance*: features are selected according to the highest relevance (dependency) to the target class c . Relevance can be interpreted such as correlation or mutual information, which defined dependencies between variables.

For discrete variables, the mutual information I of two variables x and y is based on their joint probabilistic distribution $p(x, y)$ and their probabilities $p(x)$ and $p(y)$:

$$I(x, y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}. \quad (5)$$

Searching according *maximal relevance* chooses the feature with high relevance to the target class c , so that these features have strong dependency to this class, therefore *maximal relevance* is associated with *maximal dependency*. *Maximal relevance* (maximal dependency - $maxD$) of feature set S with features x_i can be described with formula:

$$maxD(S, c), \quad D = \frac{1}{|S|^2} \sum_{x_i \in S} I(x_i; c), \quad (6)$$

where $maxD(S, c)$ is computed with the mean values of all mutual information values between individual feature x_i and corresponding c class.

Features selected by *maximal relevance* usually have rich redundancy (we can say that there is a strong dependency between features). When two features are heavily dependent on each other, their class-discriminant power would not change much if one of them will be removed. For this reasons *minimal redundancy* ($minR$) criteria is applied by formula:

$$minR(S), \quad R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j). \quad (7)$$

MRMR combines these two criterions *maximum relevance* and *minimum redundancy* by the operator $\phi(D, R)$, which optimize D and R simultaneously:

$$max\phi(D, R), \quad \phi = D - R. \quad (8)$$

MRMR feature selection framework were successfully applied to genes selection, where selected genes led to the improving of class prediction on five gene expression data sets such as Lymphoma, Lung, Leukemia, Colon, NCI [7].

It should be noticed that the combination a very effective features with another very effective features does not necessarily lead to the better feature set.

In general for feature selection algorithms, there are these advantages:

- a dimension reduction for reducing the computational cost,
- a reduction of irrelevant features (noise features) for improving the classification accuracy.

IV. DESCRIPTION OF EXPERIMENTS

Extended acoustic event database [10] used in this work involved the gun shot recordings with 463 realizations of shots from the most commonly used weapons, breaking glass recordings with 150 realizations of broken glass and approximately 53 minutes of background sounds (traffic sounds). The recordings (48 kHz, 16 bits per sample) were cut and manually labeled using Transcriber. In the testing process 46 gun shots and 13 breaking glass were used. This sounds form to acoustic event class and they are evaluated together. The rest of sounds were used in the training process, where acoustic models for shots, glass and background were created.

In the first step extraction of MFCC coefficients were done with HCopy tool. MFCC with energy coefficients (MFCC_E) created 13 dimension feature vectors and MFCC_E_D_A (energy, delta and acceleration coefficients) created 39 dimension vectors. For computing MFCC, 25 ms Hamming window was used. The extraction of ZCR, skewness and kurtosis was done in Matlab environment. This way 3 dimension feature vectors were created. First experiments were done with full set of parameters (MFCC_E, MFCC_E_D_A, moments and ZCR), then several experiments with ZCR, skewness, kurtosis and various combination of MFCC coefficients, which were chosen according to MRMR algorithm, e.g. MFCC_E features set was extended with three parameters (moments and ZCR) to the one supervector with 16 dimension (13+3).

Then the MRMR selection algorithm was applied to the searching and sorting MFCC coefficients from the best to the worst one. According the MRMR the normal order was changed to another sequence : 6, 5, 7, 2, 1, 3, 4, 8, 13, 10, 9, 11, 12, 28, 16, 27, 14, 29, 15, 26, 17, 30, 34, 21, 19, 20, 24, 31, 25, 18, 23, 32, 36, 37, 22, 33, 38, 39, 35.

As you can see, the basic coefficients from 1 to 13 are the most relevant, then follow the rest of coefficients (without significant differences between delta and acceleration coefficients). The selections of MFCC coefficients were done according to MRMR algorithm. The creation of reduced feature sets is described in the following lines.

A. Creation of reduced sets

First feature subset was created from first 13 coefficients (basic coefficients) - set A, another one - set H was created with them, where last three coefficients were removed (i.e. coefficient number 9, 11, 12). This removed group was replaced with skewness, kurtosis and ZCR, consequently feature vectors had dimension equal to 13 (10+3). Next subset - set D, consisted from MFCC_E and skewness, kurtosis and ZCR.

Another feature subset - set F was done with first 13 coefficients without 12th coefficient, then skewness, kurtosis and ZCR were added, hence feature vectors consist of 15 parameters (12+3).

Next feature subset - set G consist of first 13 coefficients, 16th coefficient, moments and ZCR. This coefficient belongs to the delta coefficients (first time derivation of basic coefficients). According to MRMR algorithm basic MFCC set with 28th coefficient, moments and ZCR was evaluated too - set K.

For creation of feature subset - set J, MFCC_E_D_A (i.e. 39 coefficients) was used, where last three coefficients (38th, 39th and 35th) were replaced by skewness, kurtosis and ZCR, consequently feature vectors had dimension equal to 39 (36+3).

Full set of MFCC_E_D_A represented set B, skewness, kurtosis and ZCR created set - C, MFCC_E_D_A plus moments and ZCR were in the feature set E. The structure of all testet combination is described bellow.

In our case 77 background files, 13 breaking glass and 46 gun shots were tested. Form one to four states HMMs up to 1024 PDFs were trained and tested. The one state models achieved the best results in comparison of two, three a four states HMMs. The maximum values (99,26 %) of ACC were reached with 4 and 8 PDFs. Therefore only results for one state HMMs up to 64 PDFs are presented in the Fig.1, where:

- set A - MFCC_E (13 coefficients),
- set B - MFCC_E_D_A (39 coefficients),
- set C - skewness, kurtosis and ZCR (3 coefficients),
- set D - MFCC_E, skewness, kurtosis and ZCR (16 coefficients),
- set E - MFCC_E_D_A, skewness, kurtosis and ZCR (42 coefficients),
- set F - MFCC_E without 12th, skewness, kurtosis and ZCR (15 coefficients),
- set G - MFCC_E plus 16th coefficient, skewness, kurtosis and ZCR (17 coefficients),
- set H - MFCC_E without 9th, 11th, 12th skewness, kurtosis and ZCR (13 coefficients),
- set I - MFCC_E_D_A, first 19 coefficients, skewness, kurtosis and ZCR (22 coefficients),
- set J - MFCC_E_D_A, without 35th, 38th, 39th coefficients, skewness, kurtosis and ZCR (39 coefficients),
- set K - MFCC_E plus 28th coefficient, skewness, kurtosis and ZCR (17 coefficients),
- set L - MFCC_E plus 16th plus 28th coefficient, skewness, kurtosis and ZCR (18 coefficients).

These feature sets were created according to the MRMR algorithm, which sorted MFCC coefficients from the most relevant to the less relevant ones.

B. Evaluation of results

The training and recognition process were done with Hidden markov model ToolKit - HTK [9]. HTK includes different program tools that were primary developed for speech recognition tasks. HResults tool evaluates the output of HTK decoder with e.g. accuracy measure that is described following:

$$ACC [\%] = \frac{N - D - S - I}{N} \times 100, \quad (9)$$

where D is the number of deletion errors, S the number of substitution errors, I the number of insertion errors, and N is the total number of labels in the reference transcription files [9]. This measure is not oriented on the particular class (acoustic events), but it informs about overall system performance. The class which is the most represented in the reference has the largest rate on the overall results.

Another evaluation approach is based on the confusion matrix (can be obtained form HResults) and several measures can be computed e.g. precision, recall, F - measure, etc. This group of measures is more suitable for representing the class

TABLE I
RESULTS OF EXPERIMENTS [%] FOR DIFFERENT FEATURE SETS

PDFs	[%]	set A	set B	set C	set D	set E	set F	set G	set H	set I	set J	set K	set L
1	ACC	17.65	60.29	75.00	77.94	70.59	77.94	79.41	78.68	78.68	65.44	80.88	81.62
	Precision	5.08	68.06	71.03	76.36	76.12	76.36	77.19	76.79	82.26	71.83	78.95	79.31
	Recall	5.08	96.08	74.55	77.78	100	77.78	78.57	78.18	96.23	100	80.36	80.70
2	ACC	35.29	50.74	57.35	88.24	69.12	88.24	89.71	85.29	83.82	66.91	89.71	93.38
	Precision	25.42	60.71	51.72	86.44	73.97	86.44	86.89	81.97	82.61	72.22	86.89	91.67
	Recall	25.42	100	54.55	89.47	96.43	89.47	89.83	84.75	100	96.30	89.83	94.83
4	ACC	61.67	55.88	46.32	94.12	52.94	94.12	94.12	98.53	72.06	58.82	94.12	94.12
	Precision	56.90	65.48	42.11	93.55	63.41	93.55	93.55	98.31	74.03	67.53	93.55	93.55
	Recall	60.00	100	39.34	100	96.30	100	100	100	100	96.30	100	100
8	ACC	48.53	61.03	46.32	99.26	75.74	99.26	99.26	88.97	96.32	76.47	99.26	89.71
	Precision	41.67	68.35	38.98	98.33	77.63	98.33	98.33	86.67	93.65	79.17	98.33	88.14
	Recall	42.37	100	42.59	100	100	100	100	88.14	100	96.61	100	88.14
16	ACC	35.29	86.03	33.82	72.06	95.59	67.65	72.06	69.85	80.88	97.06	82.35	80.15
	Precision	25.42	85.29	25.00	67.80	93.65	62.71	67.80	65.00	77.97	95.16	79.66	76.67
	Recall	25.42	100	26.97	68.97	100	63.79	68.97	66.10	80.70	100	81.03	77.97
32	ACC	51.47	95.59	28.68	56.62	97.06	56.62	58.09	59.56	69.12	86.76	59.56	62.50
	Precision	44.07	95.00	18.64	50.00	95.08	50.00	51.67	53.33	63.93	83.61	53.33	56.67
	Recall	44.83	98.28	18.64	50.85	98.31	50.85	53.54	54.24	66.10	86.44	54.24	57.63
64	ACC	50.00	66.19	22.06	56.62	72.06	55.88	56.62	56.62	67.65	68.38	55.15	51.47
	Precision	42.37	63.49	100	50.00	67.21	49.12	50.00	50.00	62.30	63.33	48.33	43.86
	Recall	43.10	68.97	41.94	50.85	69.49	49.12	50.85	50.85	64.41	65.52	49.15	43.86

dependent results, therefore precision and recall were chosen to describe the classification performance for acoustic events class. Precision is defined by the following formula:

$$Precision [\%] = \frac{TP}{TP + FP} \times 100. \quad (10)$$

Recall is defined by the following formula:

$$Recall [\%] = \frac{TP}{TP + FN} \times 100, \quad (11)$$

where *TP* - true positive is the number of items correctly labeled as belonging to the positive class, *FP* - false positive is the number of items which are incorrectly labeled as belonging to the positive class, *FN* - false negative corresponds to the items which were not labeled as belonging to the positive class but should have been.

Presented results show the influence of MRMR selection to the system performance. The bold are results of accuracy up to 98%. As was mentioned before, ACC corresponds to the overall accuracy and it pointed to the overall detection performance. The best recognition scores of ACC = 99.26% for HMM of 8 PDFs for feature sets D, F, G, K were achieved and second best results 98.53% for HMM with 4 PDFs were reached. Interesting results (above 90%) were obtained for set B (32 PDFs), set E (16, 32 PDFs), set F (4 and 8 PDFs), set G (4 and 8 PDFs), set H (4 PDFs), set I (8 PDFs), J (16 PDFs), set K (4 and 8 PDFs) and set L (2 and 4 PDFs).

The measures precision and recall point to the classification ability to distinguish the background sounds against acoustic events. We have not considered an error, which occurred in cases, where HMM classifier recognized a shot like a breaking glass or vice versa.

V. CONCLUSION

This paper presented partially framework of detecting potentially dangerous acoustic events. As a main feature extraction algorithm MFCC, skewness, kurtosis and ZCR were used. MFCC represented the standard method in the speech recognition and with the combination of other extraction methods (e.g. descriptors defined in MPEG-7, statistical moments, spectral

flux, spectral roll-off, ZCR, etc.) can effectively describe non-speech sounds. For the selection of relevant subsets of MFCC coefficients, MRMR algorithm was used. Optimal subset of MFCC consist of basic coefficients (13) plus 16th or 26th and the appropriate dimension of feature vectors was in a range from 15 to 17 parameters per frame. For the detection of acoustic events delta and acceleration coefficients seem to be less important.

In the future we would like to tune the processes of extraction, selection and combination of parameters.

ACKNOWLEDGMENT

The research presented in this paper was supported by the EU ICT Project INDECT (FP7 - 218086) (30%) and Research and Development Operational Program funded by the ERDF under the projects ITMS-26220120030 (35%) and ITMS-26220220155 (35%).

REFERENCES

- [1] Latané, B., Darley, J.M.: Bystander apathy. American Scientist, vol.57, No.2, (1969)
- [2] Latané, B., Darley, J.M.: Group inhibition of bystander intervention in emergencies. Journal of Personality and Social Psychology, vol.10, No.3, 215-221 (1968)
- [3] Mesaros, A., Heittola, T., Eronen, A., Virtanen, T.: Acoustic event detection in real life recordings. In: EUSIPCO 2010, 1267-127, Denmark (2010) ISSN: 2067-1465
- [4] Kim, H.G., Moreau, N., Sikora, T.: MPEG-7 audio and beyond: Audio content indexing and retrieval. Wiley, pp. 304 (2005) ISBN: 978-0-470-09334-4
- [5] Casey, M.: General sound classification and similarity in MPEG-7. Organised Sound, 6 (2), 153-164 (2001)
- [6] Mitrovic, D., Zeppelzauer, M., Eidenberger, H.: Analysis of the data quality of audio descriptions of environmental sounds. In: Journal of Digital Information Management, vol. 5/2, pp 48-55 (2007)
- [7] Peng, H., Long, F., Ding, Ch.: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 8, 1226-1238 (2005)
- [8] Ding, C., Peng, H.: Minimum redundancy feature selection from microarray gene expression data. In: Bioinformatics Conference, CSB 2003, 523 - 528 (2003) ISBN: 0-7695-2000-6
- [9] Young, S., et. al.: The HTK Book. Cambridge University (2009)
- [10] Pleva, M., Vozarikova, E., Dobos, L., Cizmar, A.: The joint database of audio events and backgrounds for monitoring of urban areas. In: Communications in computer and information science: Multimedia Communications, Services and Security, Vol. 149, p. 191-197 (2011)

Crawling dynamic online networks

¹*Gabriel LUKÁČ (3rd year)*

Supervisor: ²Marián MACH

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹Gabriel.Lukac@tuke.sk, ²Marian.Mach@tuke.sk

Abstract—Crawling dynamic networks is a process of collecting large amount of network data from dynamic online media. Although there has been a lot of effort on designing effective crawling strategies for web, very few was done in the field of crawling online social networks. In this paper we review two random walk algorithms for sampling graph data. In the end we propose a methodology for online sampling of large dynamic datasets with the intention to extract a representative sample of the original social network.

Keywords—complex network, crawling, sampling, random walk, Kolmogorov-Smirnov test.

I. INTRODUCTION

Crawling dynamic networks is a process of collecting large amount of online data that provide a partial overview about the structure and undergoing processes. The application of data harvested by crawling is very wide. Social sciences use it as a valuable source for studying behaviour of online communities. Marketers are seeking for influential users, companies are trying to discover, what their customers think about their products and services. But the usage of such data does not stick solely on practical application. Their use in large scale data mining is invaluable. They serve as a reference to test new methods to discover structures of communities and communication patterns inside them.

Crawling is a systematical process of uncovering structure of underlying online network. More formally, it is a process selecting a subgraph of the original graph meeting predefined quality criteria and satisfying constraints determined by technical environment. Additionally, the usage scenario directly requires, that the subset should reflect properties of the original graph. From this perspective extracting such a representative dataset is a non-trivial task.

Our paper is divided into following sections: In section II we deal with related work. Section III is devoted to basic definitions. In section III-A we provide basic description of our approach.

II. RELATED WORK

Design of crawling algorithms has been a hot topic for last ten years. A lot of effort has been put on designing effective crawling strategies for the web crawling. But very little has been done on crawling of online social networks. From the work related to ours it is necessary to mention [1] for development of freshness and age measures of data stored in crawler's local databases. Authors in [10] analyzed and designed algorithms for optimal scheduling of web crawlers.

In the area of graph sampling authors [4] have made an extensive work on analyzing a broad range of sampling algorithms and evaluated them on a large set of graph properties. A recent work of [3] is very closely related to our work. It proposes to use Metropolis-Hastings algorithm to deal with the bias of simple random walks towards high degree nodes. But from the engineering and crawler design perspective, there has been almost no work available on combining optimal scheduling and representative graph sampling. The combination of above-mentioned two challenges gives an opportunity to design an engine capable of crawling representative sample of online social networks.

III. CRAWLING ONLINE NETWORKS

A web crawler is a system for the bulk downloading of web pages [7]. Its prevalent purpose is to be used as a key information harvester of web search engines. The control algorithm of such a system is simple. Start with a queue pre-filled with a seed of URL addresses and download their content. In the next step, extract the hyperlinks present at examined web sites. Add hyperlinks that have been extracted into the queue. Iteratively continue with sequential downloading of URL addresses available in the queue and subsequently discover new hyperlinks.

Although the crawling process itself appears to be simple, there are several challenges that every system should count with, especially the large scale of data.

In addition, every crawling tool must deal with a very dynamic nature of such data. Because of the large scale of data, crawling a of the Web can last for weeks or months. By the time a Web crawler has finished its crawl and created a local copy of the online data there is a significantly high probability, that the data available online have changed. In order to characterize the quality of the local copy two freshness metrics have been proposed [1]:

- 1) **Freshness:** It can be characterized as a fraction of the local database that is up-to-date. More formally, let the freshness of a local element e_i at time t be:

$$F(e_i; t) = \begin{cases} 1 & \text{if } e_i \text{ is up-to-date at time } t \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Then the freshness of the local database S at time t is defined as

$$F(S; t) = \frac{1}{N} \sum_{i=1}^N F(e_i; t). \quad (2)$$

- 2) **Age:** A metric to estimate *how old* the local database is defined as the average of local element *ages* at time t

$A(e_i; t)$. If the local "age" of the element e_i at time t is defined as follows:

$$A(e_i; t) = \begin{cases} 0 & \text{if } e_i \text{ is up-to-date at time } t \\ t - m_i & \text{otherwise,} \end{cases} \quad (3)$$

where m_i is the modification time of the element e_i , then the overall age of local database S is defined as:

$$A(S; t) = \frac{1}{N} \sum_{i=1}^N A(e_i; t). \quad (4)$$

For details review the work [1].

Dealing with dynamic nature of data is even more crucial in crawling large scale social networks such as Facebook or Twitter. The tendency to change in such a live media resources is much more evident. In order to design a cut-of-edge crawling engine for online social networks one must take into the consideration all of the above-mentioned dynamic aspects.

A. Crawling a representative sample

In general crawling a large-scale network can be perceived as sampling a subgraph of an original graph. To implement a useful sampling strategy, the overall usage purpose should be always taken into account. In order to use data for graph mining applications, they should provide a certain level of representativeness of the original data. In order to select a representative subgraph of the original graph, several approaches have been proposed on the past (see section II for details). However, in most cases they are being used for offline sampling what makes them unsuitable for crawling dynamic networks.

In our approach, we are building on the top of work [3], which proposes to use Metropolis-Hastings Random Walk algorithm for unbiased online network sampling. We use it to sample subgraphs of various sizes (starting at 5% and ending at 100% of vertices) of the original graph. To estimate the quality of particular samples we use the Kolmogorov-Smirnov D-statistic as a metric of the similarity between the distributions of selected properties (e.g. node degree distributions, clustering coefficients distribution, etc.) between samples and the original graph. The approach of using D-statistic to compare graph sample with the original has been proposed by [4].

Our methodology of online network sampling consists of two steps.

- 1) At first we run Metropolis-Hastings Random Walk algorithm to sample a subgraph with similar topological characteristics to original graph. The main purpose of this procedure is estimation of global properties of the original graph. As it is impossible to sample the original graph completely (due to the large scale of the original graph and technical constraints of the available infrastructure), we propose an approach to detect convergence of random walk algorithm on a reasonably representative reference sample. The approach is based on comparing every two consecutive samples of different sizes by Kolmogorov-Smirnov test. When a predefined number of reasonably small changes between two consecutive property distributions occur, the algorithm declares convergence and finishes. The last sample is consequently used as a reference approximation of the original graph.

- 2) The second step is online sampling as such. The application of the data defines requirements for the crawling algorithm. The algorithm designer should incorporate into the design the freshness and age of the local database, and the algorithm should be capable of working under constraints predefined by technological environment and infrastructure. What makes the design of online sampling algorithm easier is the availability of global topological properties of the original graph that have been approximated by the reference sample in the previous step. In most cases, there is no such additional information available. Due to the limited space of this paper, we do not provide here any further information about this step.

IV. PRELIMINARY RESULTS

In this section preliminary results are given. For the evaluation purposes two different real-world networks are used. The first is a snapshot of Gnutella peer-to-peer network from August 2002 [5][8] with 6301 nodes and 20777 edges. Nodes represent hosts in the Gnutella network, directed edges connections between them. The second dataset is the High energy physics theory citation network [5][2] with 27770 papers and 352807 citations.

A. Analysis of Sampling Algorithms

In order to select appropriate approach for online sampling, two versions of Random walk algorithms have been studied. The first is the simple version of random walk, where the walk is characterized as a stochastic process with random variables X_1, X_2, \dots, X_k such that $X_1 = 0$ and $X_i + 1$ is a vertex chosen uniformly at random (with transition probability $P_{v,w}^{RW}$) from the neighbors of X_i [9]. As it has been proved that the simple random walk algorithm is biased towards high degree nodes [6], we considered to use Metropolis-Hastings algorithm to correct the bias and converge to a uniform distribution. The technique achieves the goal by using the following transition probability:

$$P_{v,w}^{MH} = \begin{cases} \frac{1}{k_v} & \text{if } w \text{ is a neighbor of } v, \\ 1 - \sum_{y \neq v} P_{v,y}^{MH} & \text{if } w = v, \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where k_v is the degree of node v .

In order to compare selected algorithms and observe the expected improvement, two experiments on both datasets have been conducted. For each sample size (starting at the level of 5%, ending at 100%) we perform 10 random walks. After that we compute the in-degree, out-degree and clustering coefficients distributions for each of the samples and compare each of them by Kolmogorov-Smirnov test to the original graph. The value of D-statistic indicates an error of sample representativeness. On Fig. 1 the evolution of D-statistic values for particular sample sizes are shown. We can make an observation that for the most of sample sizes the Metropolis-Hastings random walk generates samples of better representativeness than a simple random walk. For the High energy physics dataset the affirmation is not so obvious (see Fig. 2). However there is a clean observable tendency of Metropolis-Hasting algorithm to steadily improve values of D-statistic for each sample size. Instead of that, the simple random walk is provably much more chaotic. Although for higher sizes of samples it converges reasonable well.

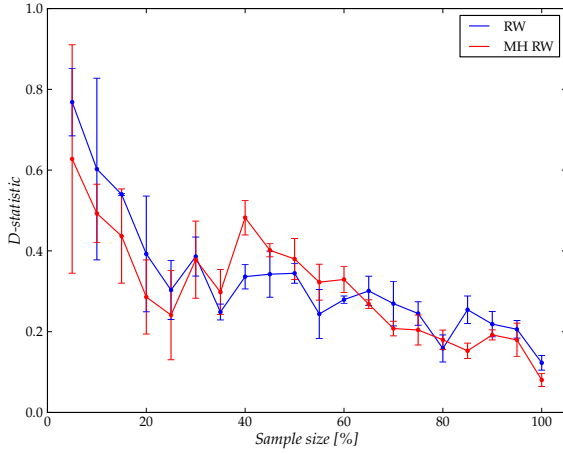


Fig. 1. Sampling of the Gnutella p2p network. The average values of D-statistic are given for every sample size for both the simple random walk (RW) and Metropolis-Hastings random walk (MH RW) algorithms. For every D-statistic value a standard deviation is given. The D-statistic values have been computed for out-degree distributions.

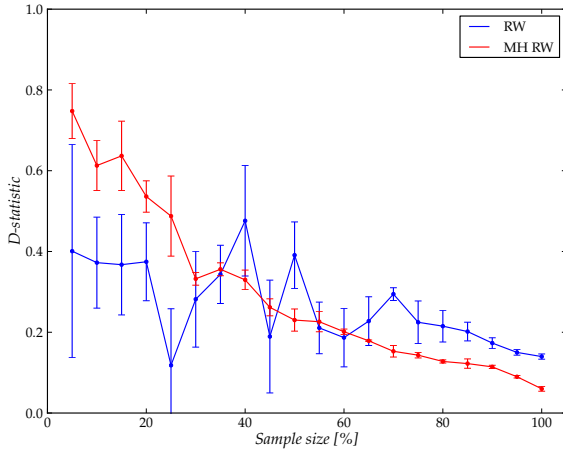


Fig. 2. Sampling of the High energy theoretical physics citation network. The average values of D-statistic are given for every sample size for both the simple random walk (RW) and Metropolis-Hastings random walk (MH RW) algorithms. For every D-statistic value a standard deviation is given. The D-statistic values have been computed for in-degree distributions.

B. Estimating Sampling Termination Clause

Since we are interested only in representative samples it is practical to propose a termination clause of the algorithm. The task itself is not trivial since to declare convergence of being representative without being able to compare with referential data is difficult. For that purpose we propose an approach based on comparing every two samples of augmentative sizes by Kolmogorov-Smirnov test. After we observe a predefined number of reasonably small changes between D-statistic values, the random walk is said to converge and the sampling algorithm terminates. The sample discovered at the time of termination is consequently used as a reference approximation of the original graph.

In order to test our approach we performed two experiments on Gnutella p2p and High energy teoretical physics datasets. As it can be seen on Fig. 3 by using Metropolis-Hastings algorithm convergence can be declared after sampling around 35% of the original graph.

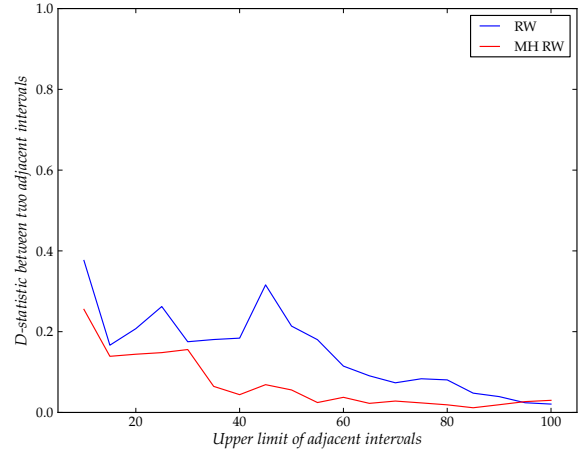


Fig. 3. Change of D-statistic values between two consecutive sample sizes provided for both - simple random walk (RW) and Metropolis-Hastings random walk (MH RW). The changes have been computed for in-degree distributions.

V. CONCLUSION

In this paper we described crawling of online networks as a challenging problem. Two experiments on two different datasets have been provided and we have shown a clear advantage of using Metropolis-Hastings random walk algorithm for sampling network data. In order to perform representative online sampling we proposed a two-step methodology where in the first step a reference sample of the original graph is obtained and the second step is devoted to the online sampling itself. To detect the convergence of reference sampling we proposed an approach bases on comparing the change of D-statistic between two samples of consecutive size levels.

ACKNOWLEDGMENT

The work presented in this paper was supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic under grant No. 1/0685/12 (50%). This work is also the result of the project implementation Development of the Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (50%).

REFERENCES

- [1] Junghoo Cho and Hector Garcia-Molina. "Synchronizing a database to improve freshness". In: *SIGMOD Rec.* 29.2 (May 2000), pp. 117–128. ISSN: 0163-5808. DOI: 10.1145/335191.335391. URL: <http://doi.acm.org/10.1145/335191.335391>.
- [2] Johannes Gehrke, Paul Ginsparg, and Jon Kleinberg. "Overview of the 2003 KDD Cup". In: *SIGKDD Explor. Newsl.* 5.2 (Dec. 2003), pp. 149–151. ISSN: 1931-0145. DOI: 10.1145/980972.980992. URL: <http://dx.doi.org/10.1145/980972.980992>.
- [3] Minas Gjoka et al. "Walking in facebook: a case study of unbiased sampling of OSNs". In: *Proceedings of the 29th conference on Information communications. INFOCOM'10*. San Diego, California, USA: IEEE Press, 2010, pp. 2498–2506. URL: <http://dl.acm.org/citation.cfm?id=1833515.1833840>.

- [4] Jure Leskovec and Christos Faloutsos. “Sampling from large graphs”. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. KDD '06. Philadelphia, PA, USA: ACM, 2006, pp. 631–636. ISBN: 1-59593-339-5. DOI: 10.1145/1150402.1150479. URL: <http://doi.acm.org/10.1145/1150402.1150479>.
- [5] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. “Graphs over time: densification laws, shrinking diameters and possible explanations”. In: *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. KDD '05. Chicago, Illinois, USA: ACM, 2005, pp. 177–187. ISBN: 1-59593-135-X. DOI: 10.1145/1081870.1081893. URL: <http://doi.acm.org/10.1145/1081870.1081893>.
- [6] L. Lovász. “Random Walks on Graphs: A Survey”. In: *Combinatorics, Paul Erdős is Eighty*. Ed. by D. Miklós, V. T. Sós, and T. Szőnyi. Vol. 2. Budapest: János Bolyai Mathematical Society, 1996, pp. 353–398.
- [7] Christopher Olston and Marc Najork. “Web Crawling”. In: *Found. Trends Inf. Retr.* 4.3 (Mar. 2010), pp. 175–246. ISSN: 1554-0669. DOI: 10.1561/1500000017. URL: <http://dx.doi.org/10.1561/1500000017>.
- [8] Matei Ripeanu, Adriana Iamnitchi, and Ian Foster. “Mapping the Gnutella Network”. In: *IEEE Internet Computing* 6.1 (Jan. 2002), pp. 50–57. ISSN: 1089-7801. DOI: 10.1109/4236.978369. URL: <http://dx.doi.org/10.1109/4236.978369>.
- [9] Wikipedia. *Random walk* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 12-March-2012]. 2012. URL: http://en.wikipedia.org/w/index.php?title=Random_walk&oldid=480770522.
- [10] J. L. Wolf et al. “Optimal crawling strategies for web search engines”. In: *Proceedings of the 11th international conference on World Wide Web*. WWW '02. Honolulu, Hawaii, USA: ACM, 2002, pp. 136–147. ISBN: 1-58113-449-5. DOI: 10.1145/511446.511465. URL: <http://doi.acm.org/10.1145/511446.511465>.

Application of 3D sensors in the design of automatic unloading mechanisms

¹Anton MOLČAN (1st year)
 Supervisor: ²Anna JADLOVSKÁ, ³Ján LIGUŠ

^{1,2,3}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹anton.molcan@tuke.sk, ²anna.jadlovaska@tuke.sk, ³jan.ligus@tuke.sk

Abstract—With 3D scanning devices and powerful libraries for computer vision recently becoming far more available on the market new options are available for practical application design. The task of cargo unloading still relies mostly on human workforce and might be a good goal for automation attempts. In this paper we describe several experiments aimed to explore the possibilities we have available in this field.

Keywords—3D sensors, automatic unloading, computer vision, image registration

I. INTRODUCTION

The main goal of process automation is to remove human labor from menial tasks at the lowest level of process control, thus eliminating the disadvantages of human workforce. When correctly designed, machines are less prone to errors, tireless and most often less costly. Also, removing humans from direct process control greatly reduces workplace accident hazards.

One of such menial tasks is the task of cargo unloading or depalletization. While automatic designs do exist for this task, most companies in this region still rely on human workforce for this task. It is the goal of this paper to describe the conditions and options available for the design of an automatic unloading mechanism.

During the last year 3D scanning devices became available on the commercial market at a low price. These devices may be used in open systems design and effectively eliminate the need for sophisticated multi camera arrays to infer spatial information from scenes. As they have means to directly measure distances, these devices require no calibration and may be mobile, therefore our design counts on such devices.

Usage of such devices is supported by the fact that several compatible open-source projects are available, containing many efficient algorithms for computer vision applications.

Robotic manipulators are quite common in industrial applications. For the purposes of this paper it is sufficient to say that enough such devices, with different properties are readily available on the market.

II. TASK DESCRIPTION

The unloading process is described as follows. A manufacturing plant creates goods, which are packaged in a cargo container and shipped to a different plant for additional

processing and/or sale. If the cargo is not packaged in special containers or by special devices and as the cargo is stirred during transit, we generally may not assume any pre-defined spatial pattern of the particles (single pieces). Every particle has a random position, which means that we face a 6 degrees of freedom (DoF) identification problem, where the 6 DoF are coordinates in an xyz reference system and rotation in the respective planes (as compared to a base model). This problem is known as 3D shape registration. Practical applications may also have several additional specifications, such as precise positioning of the unloaded parts for further processing.

As the proposed solution is supposed to replace human workforce from the process it has to meet certain requirements

- Time efficiency: the proposed solution must be less or equally time consuming than the available human labor.
- Cost efficiency: the combined cost of devices and cost of maintenance should be less than the cost of human labor over a set period of time.
- Reliability: no piece of cargo must remain in any container, if the system is unable to retrieve a piece it has to notify the operator.
- Precision: with insufficient precision there is a risk of damaging the components of the system.
- Robustness: the conditions of the operating environment should be taken into consideration.

There are also several constraints implied by the proposed measurement method (optical measurement).

- Model: every container should only contain one kind of particles and it has to be possible to obtain a finite size model (analytical or experimental) of a particle (ideally beforehand). Different particles should have minimal deviation.
- Material: the material should be optically stable. That means not specular, transparent, fluorescent or otherwise distracting. Some manipulators may also require the material to be firm or have certain magnetic properties.

It may be noted, that IR pattern based scanners are not inhibited by low quality lighting. Classic camera arrays often require additional light sources to correctly compute the scene, which means additional expenses.

III. 3D SHAPE REGISTRATION

With open-source projects such as OpenCV [1] and Point Cloud Library (PCL) [2] the effort for designing complex visual systems has been greatly reduced. For this application we shall consider feature-based registration methods and we presume that only rigid linear transformation of the object occurs.

While efficient algorithms for feature detection are available in 2D space, such as SIFT [3] and SURF [4], working in 3D extends the available options.

In [5] the author describes 3D point feature representations as being able to capture a local surface characteristic in the presence of rigid transformations, varying sampling density and noise. Examples of such features, implemented and ready for use in the PCL are Point Feature Histograms (PFH), Viewpoint Feature Histograms (VFH), Normal Aligned Radial Features (NARF), etc.

In combination with several other algorithms for image preprocessing, feature based registration algorithms should provide us with efficient and accurate means of particle localization.

For particle localization we require a model particle, in reference to which the parameters are calculated. There are many algorithms for obtaining a 3D object model, several forms of model representation, as well as the option to define one analytically or from blueprints. However analytical models are hard to obtain in most cases and while defining a model from blueprint or specification might be accurate, the difference between the statistical distribution of data in such a model and in the scan of the scene by a 3D scanning device might produce significant differences when computing features.

IV. REFERENCE SPACES

It is important to note, that as we combine several devices, we have to consider the coordinates they operate in. The sensor has its own reference space and so does the manipulator. The model stored in the system's memory is also defined within some reference space.

A global coordinate system should be defined, in which the manipulator and sensor 'understand' each other. To describe the pose of a camera in 3D space we may use a proper rigid transformation (retains point distances and handedness) matrix. Such a matrix contains information on the rotation and translation of camera coordinates compared to a default global system. Such a matrix is also the result of localization, mapping model coordinates to the global system.

Consider a 3D point $\mathbf{p}=[x, y, z]^T$ in the global reference space and another point $\mathbf{p}_c=[x_c, y_c, z_c]^T$ in the reference space of the camera (3D scanner). The relation between such points is given by

$$\mathbf{p} = [\mathbf{R} | \mathbf{t}] \cdot \begin{bmatrix} \mathbf{p}_c \\ 1 \end{bmatrix} \text{ or } \mathbf{p} = \mathbf{R} \cdot \mathbf{p}_c + \mathbf{t}, \quad (1)$$

where $\mathbf{t}=[x_p, y_p, z_p]^T$ and \mathbf{R} is a proper rotation matrix (of size 3×3 , $\mathbf{R}^T = \mathbf{R}^{-1}$ and $\det(\mathbf{R})=1$). As the transformation matrix is of size 3×4 , it may be completely determined from four pairs of non coplanar sample points. The inverse transformation may be computed as

$$\mathbf{p}_c = [\mathbf{R}^T | -\mathbf{t}] \cdot \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix} \text{ or } \mathbf{p}_c = \mathbf{R}^T \cdot \mathbf{p} - \mathbf{t}. \quad (2)$$

With static mounts it is therefore a simple matter of measurement and comparison of some default position of both

devices. Mobile mounts may require supplementary methods for coordinate comparison, e.g. color markers placed conveniently around the workplace or sensors for position measurement.

The model reference space is of crucial importance, as the computed transformations or model parameters are model specific. The point of origin for the model should be well defined, e.g. as an important point in terms of geometry (center of mass, etc.) and the relation between the point of origin and the manipulator contact point should be known.

It should also be taken into consideration, that while image preprocessing does not necessarily change the point of origin of the reference space, it should always be asserted that this is indeed so to avoid any errors.

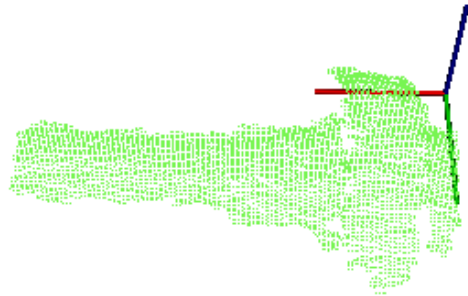


Fig. 1. A partial object model, used throughout our experiments, with point of origin and xyz axes marked on the image. The coordinates of the point of origin are selected as minimum values for each axis, so every point of the model lies in the first octant. Note that gripping point coordinates (not shown) also represent the additional translation required after particle localization.

V. EXPERIMENTAL RESULTS

In the first phase we have tried the Microsoft Kinect 3D scanner in combination with the PCL. Our model representation of preference is the point cloud representation. For the current experiments we used only a partial model. Since our particles are symmetric, this should not seriously affect the results of experiments conducted so far.



Fig. 2. Typical Kinect input data with point cloud data representation. Several particles may be seen in the image. Point cloud distances should correspond to distances in meters.

A. Layout

In our experiment we consider the following layout. The cargo container is placed on mount/conveyor A. The 3D scanner is placed above, so any mounted container is in its

effective range (ca. 1,2-3,5 m for the Kinect). Also the inside of the container should not be obstructed by anything (e.g. different mounts, walls of the container). The scanner and its mount should however not obstruct the manipulator's access to the cargo. Placement of the manipulator depends on the specific type used, but it should be placed conveniently to move particles from mount/conveyor A to output conveyor B, which may also be monitored, depending on the specification. Experiments for this paper were conducted only with a static configuration of devices. Mobile configurations are possible if required.

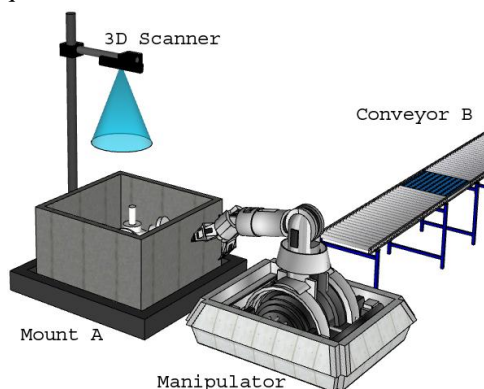


Fig. 3. Proposed general layout of our system. Static variations are shown in the figure, mobile devices and mounts are a possibility.

B. Image Preprocessing

When dealing with point clouds data volume is a concern. The Kinect device operates in 640×480 resolution for depth data. Each element is described by 3-6 values (even more in special applications) of different depth. Estimation of volume of interest (VoI) is therefore very important in terms of efficiency and time cost. Depth filtering is a basic form of lowering data volume. In our scenario a priori information may be available, such as conveyor/mount distance that may be used as a parameter, or parameter base for depth filtering. Planar segmentation is another useful concept, utilizing RANSAC [6] algorithm for planar models. The implementation available in PCL retrieves both the parameters of the planar model and the indices of the inlying points of the point cloud. This data could be important for the localization of the cargo container, for total reduction of data volume and for particle isolation.

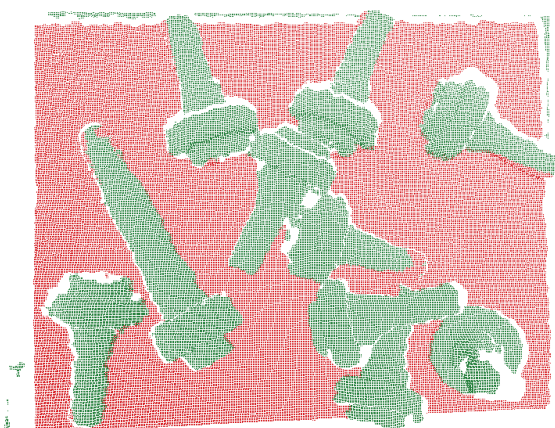


Fig. 4. Planar segmentation performed on the input data. The planar component (in red) represents a majority of points in the input cloud. The potential particles are displayed in green.

Another strategy for reducing data volume is downsampling, the process of reducing sample density. This might reduce the

precision of computations, however it greatly reduces computing time and sometimes eliminates minor deviations that make the computations unfeasible.

Clustering represents another family of algorithms that allow basic evaluation of spatial properties of point clouds. With parameters correctly set, these algorithms may either separate a much smaller VoI, or separate a particle completely. A few examples of such algorithms are Euclidean clustering and k-means.

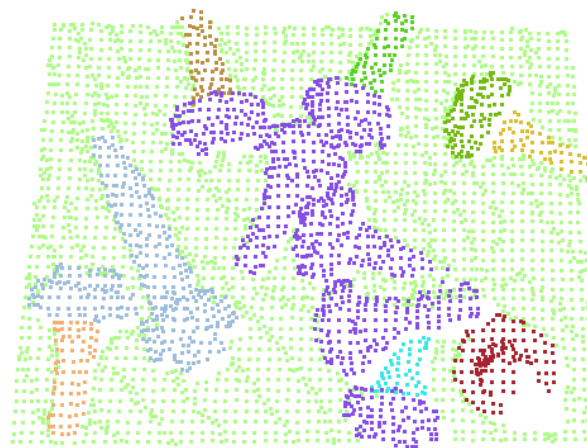


Fig. 5. Euclidean clustering applied on a downsampled point cloud. Particle positions are hinted by cluster positions. The main cluster contains several particles, but still provides a significant reduction to the VoI.

C. Particle Localization

The main point of these experiments was to determine whether the proposed combination of hard and software would provide us with usable means for particle localization. Using feature descriptors and sample consensus algorithms on preprocessed data we searched for a transformation matrix as described in (1) that would map the model to the target cloud with as many inliers as possible. We used the partial model shown in Fig. 1. A partial object model, used throughout our experiments, with point of origin and xyz axes marked on the image. The coordinates of the point of origin are selected as minimum values for each axis, so every point of the model lies in the first octant. Note that gripping point coordinates (not shown) also represent the additional translation required after particle localization.. The results are shown in Fig. 6 and Fig. 7.



Fig. 6. Result of particle localization. The transformed model is shown in blue. The model particle and one of the scanned particles overlap, thus the result is satisfactory.

```

Best fitness score: 0.000011

R = | -0.448 -0.804 0.391 |
    | 0.894 -0.393 0.216 |
    | -0.020 0.446 0.895 |

t = < -0.288, -0.103, 0.191 >

374677

```

Fig. 7. Reproduction of command line output. The **R** matrix and **t** vector correspond to (1). The fitness score is satisfactory. The bottom line shows runtime in milliseconds, which is not satisfactory.

VI. CONCLUSION

We have tested means that have recently become far more available for the task of unloading automation. The results reached and presented in this paper hint that the task is possible.

The greatest problem so far is the colossal runtime of over 6 minutes. However the proposed software was tested on a low end machine under highly nonoptimal conditions. A dedicated machine might prove significant improvement. Also programming methods such as multi-core computing and GPU utilization were not implemented in the current phase.

The conclusion of this paper is that further work needs to be conducted but it should not be futile.

ACKNOWLEDGMENT

This contribution is the results of the Vega project implementation: Dynamic hybrid architectures of the multiagent network control systems (No. 1/0286/11), supported by the Scientific Grant Agency of Slovak Republic.

REFERENCES

- [1] G. Bradski, A. Kaehler, "Learning OpenCV: Computer Vision with the OpenCV Library". Sebastopol, CA: O'Reilly Media, Inc., 2008, pp. 1-15.
- [2] R. B. Rusu, S. Cousins, *3D is here: Point Cloud Library (PCL)*, in Proceedings of the International Conference on Robotics and Automation, 2011, Shanghai, China, 2011.
- [3] D.G. Lowe, "Object recognition from local scale-invariant features", in Proceedings of the International Conference on Computer Vision 2, pp. 1150–1157.
- [4] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, "SURF: Speeded Up Robust Features", in Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, 2008, pp. 346-359.
- [5] R.B.Rusu, *Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments*. München, DE: Technische Universität München, 2009, pp. 37-96.
- [6] M.A.Fischler, R.C.Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography", in Communications of the ACM (CACM) Vol. 24, No. 6, 1981, pp. 381-395

Stratified Framework for DSL Development

¹*Sergej CHODAREV (3rd year)*

Supervisor: ²Ján KOLLÁR

^{1,2}Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹sergej.chodarev@tuke.sk, ²jan.kollar@tuke.sk

Abstract—Definition of specialized languages for solving complex programming problems can greatly help in solving these problems. Development and usage of domain-specific languages requires proper supporting tools, that would allow to reuse language artifacts and even parts of a language. A new framework for DSL development is proposed in this paper. It is designed as a set of layers based on common generic syntax. Common syntax simplifies languages development and support at a cost of decreased flexibility of notations. Other layers provide increasing support for language definition based on the concepts familiar from general-purpose languages. The framework also allows to modularize language definition to make it possible to reuse parts of a language.

Keywords—Domain-specific language, generalized syntax, language schema, reusability, stratified design.

I. INTRODUCTION

Programming language is probably the most important tool of a programmer. It is actually not only a tool, but a working medium, and as such it shapes resulting program [1]. Programming language provides ways to describe the program and to structure that description.

The most important way of structuring a program is abstraction. Abstraction is a process in which one can introduce a new concept that encapsulates some part of program logic.

High-level programming languages have a lot of abstractions built in. Besides of that, they provide ways for programmer to introduce new abstractions, like functions, modules, or classes. This allows to extend the languages — introduce new concepts and operations, not originally presented in the language but constructed from existing concepts.

II. DOMAIN-SPECIFIC LANGUAGES

In some situations, mechanisms that general-purpose language (GPL) provides for creating new abstractions, are not satisfactory. They usually allow only to introduce new terms into existing language. But to clearly describe solution of some problems, it is needed to create a new language and thus use metalinguistic abstraction [2], [3].

This leads to creation of a programming language specially suited for solving problems in a specific domain. Such languages are called *domain-specific languages (DSL)* [4], [5].

Several patterns exist for design and development of domain-specific languages [6], [7]. Usually they are divided into three main groups [5]:

- external DSLs,
- internal DSLs,
- language workbenches.

External DSLs are developed as separate languages, different from the language used for development of the rest of the system. Because of this, their development is often based on the same techniques and tools as development of general-purpose languages.

Internal DSLs use a general-purpose language as a base. They use syntax of the host language, but in some specific way, so a programmer can see it as a separate language. They violate conventions of the host language and give unusual meaning to language constructs.

Language workbenches are integrated environments for development and usage of domain-specific languages. Usually they use projectional editing [8]. This means, that primary representation of a program is not a text, but an internal form, that is projected to some textual or graphical representation and can be edited using commands of development environment.

Despite existing techniques and tools, domain-specific languages still did not gain widespread usage. All mentioned types of DSLs have disadvantages. External DSLs require to develop custom language processor and other supporting tools while internal DSLs can be limited by inappropriate host language. Language workbenches solve a lot of these problems, but integration with other tools that expect textual form of programs (most notably revision control tools) is problematic.

A concept of the framework for development and processing of domain-specific languages based on textual syntax is presented in this paper. It tries to integrate positive properties of different types of DSL development. Flexibility of the framework is assured by stratified design, where higher layers provide convenient tools for common cases, while lower layers expose greater flexibility.

III. BASIC PRINCIPLES OF PROPOSED FRAMEWORK

The main goal of proposed framework is to simplify development of practically useful DSLs and allow their convenient usage. Key features that should make it possible are:

- common generic host language,
- modularized language definition (language libraries),
- integration with general-purpose language and tools.

Generic host language provides a generalized syntax that can be used as a common substrate for DSLs. The host language itself does not have its own semantics, it only provides a basic notation for values and structures. We can call it skeleton syntax, because it defines only shapes for structuring a code instead of concrete language constructs [9].

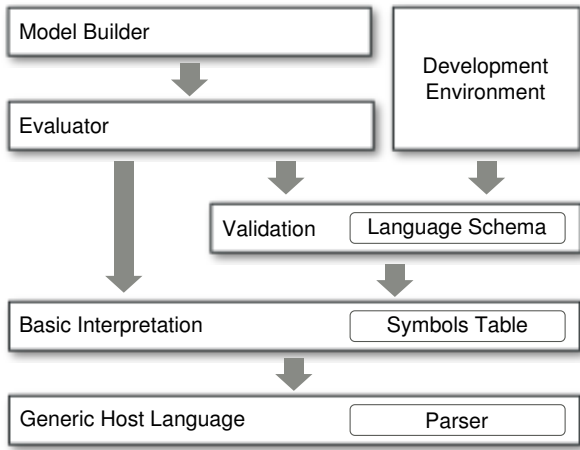


Fig. 1. Layers of proposed DSL development framework.

Usage of common syntax allows to simplify language design and development as it eliminates a need to define details of concrete syntax and to develop a parser. It also simplifies a task of language composition and tools support. This is similar to internal DSL, but in contrast to usage of existing general-purpose language as a host language, skeleton syntax allows to create a languages that does not include general-purpose constructs.

On the other hand, common syntax restricts concrete notations, that can be used by DSLs. While this may be unacceptable for some languages, a lot of DSLs can be defined on host syntax providing it is flexible enough, as is clearly shown by internal DSLs.

Modularization of language definition is required to allow reuse of language features between several DSLs [10]. A lot of languages share several common concepts and structures, so it would be a great benefit if these features could be reused in a form of language modules or libraries. This is especially important for DSLs, since a lot of them are only used in limited number of systems and therefore their development costs need to be very low.

DSLs need to be deeply *integrated with GPL and other tools* used in development of the system (like IDE). For this reason, GPL is used to describe behavior of each element of developed language. In our case implementation GPL is Java. The framework provides an Java API for manipulating programs code and define their semantics.

The main contributions of this work is a proposal of language processing framework, that would allow a new approach to the language definition based on a generic syntax and methods similar to definition of functions in general-purpose languages but extended with meta-programming capabilities.

IV. LAYERS OF DSL PROCESSING

You can see the structure of proposed framework on Fig. 1. Each layer makes development more simple and convenient for most common cases, but introduce some restrictions. However, higher layers do not cover lower ones completely. Instead, they allow to access them when greater flexibility is required.

The host language provides a parser for common syntax used by all languages. The basic interpretation layer allows to traverse a syntax tree produced by the parser and process it. The validation system and the evaluator allow to define

a language and its processing functions on a higher level, while the model builder allows to describe . Additionally development environment can provide editing support based on language schema.

A. Generic Host Language

The basic layer of proposed framework is *the generic host language* [11]. The only purpose of this language is to provide standardized concrete syntax for DSLs. It has a simple LL-class grammar, that defines notation for basic data values, identifiers and program structure elements. The number of language elements is small (see Fig. 2) and they are generic, so it is possible to use them as building blocks for a guest language.

Proposed generic language contains syntax for data values: numbers, strings, and booleans. Together with symbols (identifiers) they form atomic elements of the language.

Several types of structures are supported, that can be build from atomic elements and other structures. They are divided into data structures and control structures. Data structures are lists and maps (list of key-value pairs). They are intended for structuring static data elements.

Control structures allow to organize the program code itself. There are two types of them:

- 1) *Combination* (or horizontal structure) — a sequence of language elements separated by spaces, optionally enclosed in parentheses.
- 2) *Block* (or vertical structure) — a sequence of expressions separated by new line character and with the same indentation level.

Proposed syntax for the general host language is described in [11]. It is based on the notations used in general-purpose languages. Most influence comes from Python and Haskell languages, that use blocks structure defined by indentation. An example script using proposed syntax is provided on Fig. 3.

B. Basic Interpretation

The next layer of the framework is the basic interpretation system. It provides a simple code traversal API for implementation language, that can be used as a basis for interpretation

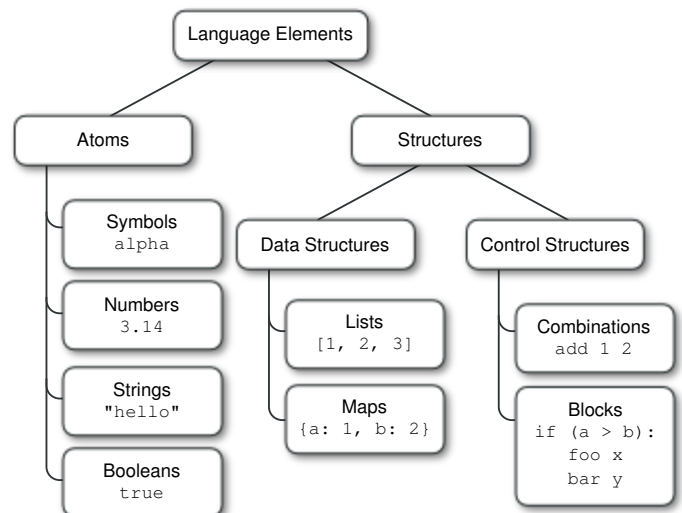


Fig. 2. Elements of generic host language with examples of concrete syntax.


```
computer:
  processor {type: alpha, cores: 2}
  disk {size: 80, speed: 7200,
        interface: ATA133}
  disk {size: 320}
```

Fig. 3. An example script in a DSL based on proposed syntax. It shows a simple language for defining properties of a computer and its components.

of DSL program. Interpretation there includes also program validation and other operations on a program code.

The interpretation system maintains a symbols table, that stores values bound to symbols in a program. This table also contains references to handler objects for interpreting program elements. If the interpreter is called on a part of syntax tree, like expression or block, it invokes a handler for the first symbol in a structure with other nodes as arguments of the handler. If it is needed, the handler can recursively apply the interpreter to arguments.

The symbol table has a stack based structure with several levels corresponding to name scopes of interpreted program. A handler of a block can append new scope level to the table while interpreting contents of the block.

C. Validation and Language Schema

Languages based on proposed framework will share the same host syntax. However, they would have different control structures and operations. To provide automatized program validation, a language definition mechanism needs to be provided. This definition mechanism is called *language schema* and it describes elements provided by a defined language.

The validator based on the schema is build on top of the basic interpretation system by providing special handlers for checking program elements. This will a) simplify development of program evaluator by removing a need to manually check all constraints, and b) allow tools (like development environment) to check program validity based on language definition.

Schema need to contain a list of language elements and describe their properties. These include:

- a symbol bound to the element,
- number and types of parameters,
- a type of return value,
- new symbols bound by the element execution,
- if element can introduce a block,
- if that block defines a new scope.

Basic properties defined by a schema are similar to the type of a function. However, language elements are much more versatile than functions, so additional properties are added for specifying if an element binds values to some new symbols and if it can introduce a block. Elements that can appear inside a block can also be restricted based on their type. This allows to define quite complex elements and constraints of their usage. An example of element declaration is given on Fig. 4. It describes element *const* with two parameters that defines a new symbol with a name of first parameter and type *Number*.

Language definition can be divided into several *modules*. Each module has separate schema. A master schema of a language specifies which modules are included into the language, together with renaming rules for a case of name clash. Elements from different modules can be freely intermixed inside a program while type rules are fulfilled.

```
element const:
  doc "Define a numeric constant."
  params [sym Symbol, val Number]
  defines 1 Number
```

Fig. 4. An example schema defining one languages element.

Types are used there to restrict composition of program elements and not to describe data structures. Because of that, type system needs to allow usage of simple types identified by name, parametric types, and definition of subtyping rules.

Schema also allows to leave-out precise description for some elements so custom validation code could be provided.

D. Evaluation

Based on the basic interpretation system and validation system, the high-level evaluation system is built. It is also based on handlers associated with program symbols and written in general-purpose language. However, these handlers operate on higher level. Their parameters are automatically evaluated and new scope levels for blocks are automatically introduced. This makes development of language element handlers very similar to definition of a convenient function. Actually, handlers on this level are implemented as specially annotated methods.

Unlike classic methods, handlers at can still access a syntax tree of their parameters. They can receive parameters in different forms and which one is actually used is specified in language schema. While general-purpose languages usually differentiate two kinds of values: r-values and l-values [12]; handlers on this layer support three kinds:

- 1) *r-value* — data value as a result of evaluation of program part,
- 2) *l-value* — data storage object assigned to a symbol,
- 3) *s-value* — syntax tree representation of program part.

First one, r-value, corresponds to convenient call-by-value arguments passing. It is useful for simple cases, where handler needs to process values produced by handlers of its arguments. L-value corresponds to call-by-reference arguments passing and allows to modify a value stored inside symbol table.

S-value (from “symbolic value”) is the most flexible way of arguments passing. The best analogy from other programming languages is a syntactic macro. In contrast to macros, handlers accepting s-values are not run prior to program evaluation and do not output a code that replaces macro call. Such handlers are executed in the process of evaluation in the same way as other handlers, but arguments that are passed as s-values are not evaluated. That arguments are passed as syntax trees and can be processed as needed.

S-values arguments provide a way to go down to basic interpretation layer if processing of the program requires greater flexibility. Other ways of argument passing provide a high-level interface.

E. Model Builder

In many cases direct evaluation of a DSL program is not the best option. Instead, a model of a program is constructed and processed in some way. Since it is quite common case, model building system is included in the framework. It provides a language for declarative description of rules for model population by DSL program. This description is then translated into evaluator, that will produce populated model.

F. Development Environment

DSL development framework does not only cover language creation, but also its usage. A basic tool in this area is an editor for DSL programs. The editor can benefit from common host syntax and even more from language definition in a schema. Based on that, it will provide at least these features:

- syntax highlighting,
- marking of syntax and validation errors directly in edited code,
- intelligent autocompletion based on a type of expected expression,
- integrated help system for language elements.

The editor can be integrated into existing Integrated Development Environment used for development in the general-purpose language. In this way it will be also better integrated into overall workflow of programmers.

G. Other layers

The framework can provide additional helper layers. One of them may be a *syntax preprocessor* that could provide custom notations not allowed by the generic syntax. This preprocessor would translate custom notations into standard generic syntax, so the rest of layers would work as usual.

Other layer may be a *code generation system*. It will help in the task of transforming program model into code in other language.

V. RELATED WORKS

Domain-specific languages is an area of intensive research [4]. Several tools are developed to simplify development of DSLs. Complex environments for language definition and usage are provided by language workbenches, for example MPS [13] or Intentional Software [14]. But their usage of projectional editing makes it complicated to integrate them with other tools.

Other tools try to provide environment similar to language workbenches for languages based on textual syntax, for example Spoofox [15] or Xtext [16]. In contrast to the proposed framework, these tools require to define custom syntax for a language. This provides greater flexibility, but on the other hand makes language definition more complex.

XML [17] and related technologies and tools create a whole framework for defining custom languages based on generic syntax. However, while XML is good for inter-program communication, its noisy syntax makes it hard for humans to read and write [18]. An exception is a usage of XML as a markup language, where XML elements are used only to mark some parts of mainly textual document.

Another attempt to define generic syntax for DSLs is presented in [19], but it does not propose processing framework and support tools.

VI. CONCLUSION

Proposed concept of the framework for DSL development is an attempt to provide developers a tool, that will make a process of language development as simple as development of library. It combines techniques familiar to the programmer from development in general-purpose languages with flexibility of manipulating with program code. It tries to allow simple composition of languages from reusable modules. It

also proposes quite powerful development environment with assistance features derived from language definition.

Usage of common generic host language for all developed DSLs is responsible for some of the advantages of proposed framework, but it also restricts notations used by languages. In the same time, common syntax not only simplifies development and languages composition, but also allows users of DSLs reuse their knowledge in cases if several languages are used.

A prototype implementation was developed to evaluate proposed concept. Development of several DSLs in it showed plausible results allowing to use techniques similar to internal DSLs development but with more convenient notation.

ACKNOWLEDGMENT

This work was supported by SRDA Project of Slovak-Slovenian Research and Development No. SK-SI-0003-10 “Language Patterns in Domain-specific Languages Evolution.”

REFERENCES

- [1] R. P. Gabriel. (2003) The Art of Lisp & Writing. [Online]. Available: <http://www.dreamsongs.com/ArtOfLisp.html>
- [2] M. P. Ward, “Language-oriented programming,” *Software - Concepts and Tools*, vol. 15, no. 4, pp. 147–161, 1994.
- [3] H. Abelson and G. J. Sussman, *Structure and Interpretation of Computer Programs*, 2nd ed., ser. MIT Electrical Engineering and Computer Science. The MIT Press, 1996.
- [4] A. van Deursen, P. Klint, and J. Visser, “Domain-specific languages: an annotated bibliography,” *SIGPLAN Not.*, vol. 35, pp. 26–36, June 2000.
- [5] M. Fowler, *Domain Specific Languages*. Addison-Wesley Professional, 2010.
- [6] 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.
- [7] D. Spinellis, “Notable design patterns for domain-specific languages,” *Journal of Systems and Software*, vol. 56, no. 1, pp. 91–99, 2001.
- [8] M. Fowler. (2005) Language workbenches: The killer-app for domain specific languages? [Online]. Available: <http://martinfowler.com/articles/languageWorkbench.html>
- [9] J. Bachrach and K. Playford. (1999) D-expressions: Lisp power, dylan style. [Online]. Available: <http://people.csail.mit.edu/jrb/Projects/dexprs.pdf>
- [10] M. Völter, “From programming to modeling - and back again,” *IEEE Software*, vol. 28, pp. 20–25, 2011.
- [11] S. Chodarev, M. Vagač, and J. Kollár, “Proposal of generic syntax for domain-specific languages,” *Journal of Computer Science and Control Systems*, no. 1, pp. 33–38, 2011.
- [12] C. Strachey, “Fundamental concepts in programming languages,” *Higher Order Symbol. Comput.*, vol. 13, pp. 11–49, April 2000.
- [13] S. Dmitriev. (2004, November) Language oriented programming: The next programming paradigm. [Online]. Available: http://www.jetbrains.com/mps/docs/Language_Oriented_Programming.pdf
- [14] C. Simonyi, M. Christerson, and S. Clifford, “Intentional software,” in *OOPSLA '06: Proceedings of the 21st annual ACM SIGPLAN conference on Object-oriented programming systems, languages, and applications*. New York, NY, USA: ACM, 2006, pp. 451–464.
- [15] L. C. L. Kats and E. Visser, “The Spoofox language workbench. Rules for declarative specification of languages and IDEs,” in *Proceedings of the 25th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications, OOPSLA 2010, October 17-21, 2010, Reno, NV, USA*, M. Rinard, Ed., 2010, pp. 444–463.
- [16] M. Eysholdt and H. Behrens, “Xtext: implement your language faster than the quick and dirty way,” in *Proceedings of the ACM international conference companion on Object oriented programming systems languages and applications companion*, ser. SPLASH '10. New York, NY, USA: ACM, 2010, pp. 307–309.
- [17] T. Bray, J. Paoli, C. Sperberg-McQueen, E. Maler, and F. Yergeau, *Extensible Markup Language (XML) 1.0*, 1998.
- [18] T. Parr. (2001, 8) Humans should not have to grok XML. [Online]. Available: <http://www-106.ibm.com/developerworks/xml/library/x-sbxxml.html>
- [19] J. Falcon and W. R. Cook, “Gel: A generic extensible language,” in *Proceedings of the IFIP TC 2 Working Conference on Domain-Specific Languages*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 58–77.

A review of data mining applications in manufacturing

¹Alexandra LUKÁČOVÁ (1st year)

Supervisor: ²Ján PARALIČ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹alexandra.lukacova@tuke.sk, ²jan.paralic@tuke.sk

Abstract—The pressure from high competition in the manufacturing sector has created a necessity to enhance productivity, to understand the process and to find the weaknesses of the process of production. In fulfilling of all these demands knowledge induction from data can be helpful. Data mining makes it possible to predict and improve future system performance and is essential in building competitive advantage also in production sector. A typical example of knowledge related to this sector may be rules or relationships for detecting failures of manufacturing process systems or for finding defects of the quality of the products. This article presents an overview of current data mining applications in production and presents a small experiment performed on production data from a chemical company.

Keywords—data mining, manufacturing data, decision tree, quality prediction

I. INTRODUCTION

It does not matter what is the type of industry or what is the size of enterprise, each organization is constantly trying to gain a competitive advantage over their competitors. Therefore especially in recent years, in period of significant information technology advancements, when organizations are able to collect and to store huge amounts of data, one of the ways to win this fight, is to discover the most interesting knowledge usable for production improvements.

In this situation is typical to use data-oriented knowledge discovery techniques (KDD – Knowledge Discovery in Databases, or currently often named Business Intelligence or Business Analytics), which represent an iterative semi-automatic process of discovery and verification of the patterns in a large amount of data, using pattern recognition technologies and statistical and mathematical techniques [1]. Data mining, method associated with KDD, was originally used mainly as a solution for marketing departments to achieve increased sales, but now is becoming more widespread and is successfully used also in other fields such as finance, economics and business problems.

However, the above mentioned areas are not all. The pressure from high competition in the manufacturing industry has created a necessity to continuous enhancement of productivity as well. Manufacturing enterprises have typically lots of data collected in huge quantities. Unfortunately, there has not been similar research interest and activity in the manufacturing domain, despite of the potential benefits. This

is due to poor accessibility of data and the necessity of communication with domain experts. Extracted knowledge should be understandable and useful for the manufacturing process.

A typical example of knowledge, the most valuable asset, related to this sector may be rules or relationships for detecting failures of manufacturing process systems or for finding reasons for defects of the quality of the products. This study provides some interesting works of data mining applications in manufacturing and shows first analyzes of experiment performed on data from chemical company.

A. Related work

The use of data mining in manufacturing began in the 1990s. Irani et al. [2] generalized an ID3 algorithm that predicted the outcome of future experiments under various, more general conditions. Their systems were used successfully in various semiconductor manufacturing applications in both diagnosis and process modeling.

The basic idea of Rokach and Maimon [3] by solving classification problems was to decompose the original set of features into several subsets, build a decision tree for each projection, and then combine them. It was tested on three real-life datasets. One limitation with the suggested algorithm was that it had no backtracking capabilities (for instance, removing a single feature from a subset or removing an entire subset).

Bakir [4] has tried in his study to identify the most influential attributes which have the important impact on causing defects on the items produced by a casting company. The target attribute was a representation of the type of error which occurred. This attribute could gain three different values 0, 1, and 2 based on the type of error. He used the methods of decision trees (CART, C5.0 and CHAID) and regression analysis (logistic regression) to classify these errors. Based on the results, the decision tree approach is a promising technique. Estimated accuracy for the final model was found to be 91,93% for the testing data. The generated decision tree model found nine process variables, which were presented to the quality team of the company. Some of the variables were for them meaningful, whereas some others were unexpected and interesting.

Ciflikli and Kahya-Ozyirmidokuz [5] in their study examined the improvement of manufacturing process via data mining techniques. The SPSS Clementine tool was used. The

aim was to detect and isolate machine breakdowns in carpet manufacturing and propose a C4.5 decision tree model. The difficulty of their study was that they had too many different categories of the qualitative breakdown reason attribute. In preprocessing phase they used data cleaning for incorrect attribute values. The records with missing values were completed with average values. In addition, they used attribute relevance analysis with information gain technique to reduce the number of the categorical variables to simplify the analysis. Ten-fold cross-validation accuracy evaluation was used to train and test the data matrix. The accuracy ratio of the final model was 72,811% and it has been recognized that some patterns which was found to be interesting for the process, were identified.

II. EXPERIMENT

In this section an original experiment will be presented. We performed knowledge discovery on real data from a chemical company. We used one of the most often implemented methodologies - CRISP-DM [7]. Therefore our experiment will be further described using this methodology in its particular steps.

A. Business understanding

Chemical company, that provided us with data, is engaged in the manufacturing of industrial yarns and one of their quality goals is to reduce the percentage of producing inferior fibers. As first goal we indentified the ability to understand main reasons why inferior fibers appear. For this purpose we would like to build a classification model in such a form, which would enable the main factors influencing reduced quality of produced fibers.

Some difficulties in this kind of study arise from the fact that a measure of quality can only be evaluated “out-of-process”, resulting in losses, because there is no alternative to removing defective parts from the production line. Therefore, it is helpful to incorporate machine learning methods to predict quality from some measured variables. In this section we will summarize our initial analysis to solve this problem.

B. Data understanding

Before preprocessing data it is essential to have an overall picture of our data and to highlight the relevant features of each attribute contained in a dataset, using graphical methods and calculating summary statistics, and to identify the intensity of the underlying relationships among the attributes [6]. The whole data set contains 240 observations from four month period and is characterized by noise and missing values. There are 15 attributes. These attributes (except ID and output attribute) can be divided into two categories. First group of attributes (X1-X6) represents characteristics of the polymer that are measured before the beginning of the process, second group of attributes (X7-X11) is measured after the post-condensation of the polymer. Target attribute stands for the percentage of first quality fibers. This information is an output of the testing process performed in the company. All variables are continuous.

C. Data preprocessing

In many projects, getting the data ready for analysis is the most time-consuming step in the process. It is important because low-quality data would lead to low-quality mining results [6]. There are a number of data preprocessing techniques.

At the very beginning, we noticed that two attributes contained duplicated values, due to control measuring. Because of the same influence on target (class) attribute they were eliminated. After that, every observation with missing value in class attribute was discarded (altogether 21 records). Records with missing values in predictor attributes were completed with the average value of the records with the same quality (there were 3 missing values in X3 attribute).

Data reduction leads to a reduced representation of the original dataset that is much smaller in volume. That is, mining on the reduced dataset produces the same (or almost the same) analytical results. There are a number of strategies for data reduction. In our research we used attribute subset selection, what means removing irrelevant attributes through correlation analysis [6]. It influenced three attributes (X6, X10 and X11). Before eliminating them we consulted our findings with expert. He confirmed these three attributes are subsets of X5 and X9. Relationships between them can be seen in Fig. 1, where scatter plot matrix visualizes the nature and intensity of the pair wise relationships in a single chart.

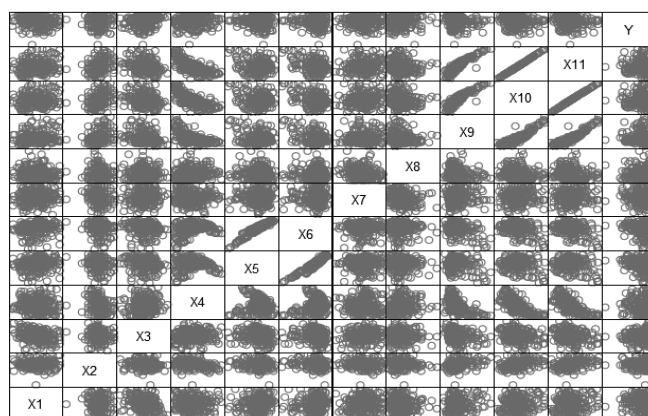


Fig. 1. Scatter plot matrix of all available attributes

We also controlled, if there are not records with the same ID. Since there was one case, we deleted these two records. This type of error could appear due to extracting data from multiple tables.

D. Application of decision tree algorithms and results

Despite the fact the target attribute is continuous; we decided to solve this problem by classification. The company's goal is to get as high quality of industrial fibers as possible, preferably above 92,2%. Therefore we divided values of output attribute into two bins as it can be seen in Fig. 2. First bin contained values greater than 92,2% (class 1) and second the others (i.e. on the contrary, lower than or equal to 92,2% - class 2). To create a classification model, using the method of decision trees, SPSS Clementine tool was utilized, which is one of the most widely used data mining software. This software supports several types of classification algorithms producing decision trees such as Quest, CHAID

and C5.0. We tried all of them, but with C5.0 we reached the best results.

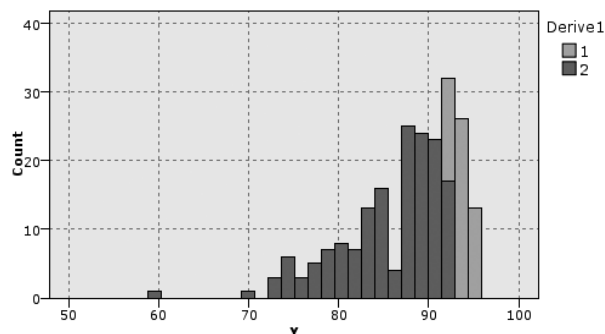


Fig. 2. Histogram of target attribute values, color resolution indicates their division into two bins

Separating data into training and testing sets is an important part of evaluating data mining models. Our training data represented 70% of full dataset. The rest was used for testing. Without any settings in the C5.0 algorithm we got interesting results. Due to small amount of observations with class 1 (23,96%), the decision tree model was not able to learn sufficiently to classify it. We can see it in Fig. 3.

Results for output field Derive1

Comparing \$C-Derive1 with Derive1

'Partition'	1_Training		2_Testing	
Correct	128	87,07%	49	70%
Wrong	19	12,93%	21	30%
Total	147		70	

Coincidence Matrix for \$C-Derive1 (rows show actuals)

'Partition' = 1_Training		1	2
1		11	19
2		0	117

'Partition' = 2_Testing		1	2
1		4	18
2		3	45

Fig. 3. Classification results, without balancing the target classes

This kind of problem we tried to solve by balancing records. It was used factor 3,7 for class 1 - records with high quality. As it can be seen in Fig. 4, the model was then able to better classify both classes, but classification results on the testing set were still not very successful. On the one hand, model correctly classified records with quality lower than 92,2%, on the other hand, it had problem to recognize the half of records with quality higher than 92,2%.

Results for output field Derive1

Comparing \$C-Derive1 with Derive1

'Partition'	1_Training		2_Testing	
Correct	142	96,6%	52	74,29%
Wrong	5	3,4%	18	25,71%
Total	147		70	

Coincidence Matrix for \$C-Derive1 (rows show actuals)

'Partition' = 1_Training		1	2
1		30	0
2		5	112

'Partition' = 2_Testing		1	2
1		11	11
2		7	41

Fig. 4. Classification results, with balance

After these findings we decided to solve this issue by next consultations with domain experts about provision of other extra inputs related to the processes that could help to improve these results. But he could not provide us with more attributes, which influence target. He explained that the outcome quality is affected by many immeasurable inputs. This is why he was

satisfied with results as well as with overall assessment efficiency of model at level of 74,29 %.

III. CONCLUSION

Manufacturing organizations are mostly interested in discovering past patterns in their production processes to predict their future behavior. In this way, data mining which is a rapidly growing interdisciplinary subject area is becoming increasingly important in manufacturing.

The aim of this study was to point out the need of applying Knowledge Discovery process in manufacturing data and by the review of some practical applications to describe their benefits.

In this paper, we have presented first analyzes on fabric data which is collected from a chemical company engaged in manufacturing of industrial yarns in Slovakia. One of their quality goals is to reduce the percentage of producing inferior fibers. Preprocessing techniques were used to improve the quality of data. In first phase of solution of the above mentioned problem we used classification model in form of decision trees in order to better understand the reasons of reduced production quality of fibers. Output values are divided into two bins. Classification results on the testing set are not very successful. On the one hand, model correctly classified records with quality lower than 92,2%, on the other hand, it had problem to recognize records with quality higher than 92,2%. This issue needed other consultations with domain experts about provision of other extra inputs related to the processes. However we came to conclusion the outcome quality is affected by many immeasurable inputs and results as well as overall assessment efficiency of model at level of 74,29 % is sufficiently satisfied.

ACKNOWLEDGMENT

The work presented in this paper was partially supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic under grant No. 1/1147/12 (30%); the Cultural Education Grant Agency under grant No. 065TUK-4/2011 (30%). This work is also the result of the project implementation Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS 26220120030) supported by the Research & Development Operational Program funded by the ERDF (40%).

REFERENCES

- [1] H. Maki, Y. Teranishi "Development of automated data mining system for quality control in manufacturing" in *Data Warehousing and Knowledge Discovery*, Springer Berlin / Heidelberg, 2001, pp. 93-100.
- [2] K.B. Irani, J. Cheng, U.M. Fayyad, et al., "Applying machine learning to semiconductor manufacturing", *IEEE Expert* 8 (1) (1993) 41-47.
- [3] L. Rokach and O. Maimon, "Data mining for improving the quality of manufacturing: a feature set decomposition approach", *Journal of Intelligent Manufacturing*, Kluwer Academic Publishers, Dordrecht, Netherlands, 17(3), 2006, pp. 285-299.
- [4] Bakır, B., Batmaz, I., Gunturkun, F. A., Ipekci, I. A., Koksall, G., & Ozdemirel. "Defect cause modeling with decision tree and regression analysis." In *Proceedings of XVII international conference on computer and information science and engineering*, 8-10 December 2006. Cairo: World Enformatika Society (Vol. 16, pp. 266-269).

- [5] Ciflikli C. "Implementing a data mining solution for enhancing carpet manufacturing productivity." In *Knowledge-based systems*, vol. 23, 2010, 783-788.
- [6] J. Han. "Data mining: Concepts and techniques." Elsevier Inc., 2012.
- [7] Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T. "CRISP-DM 1.0 Step by step data mining guide" 2000 SPSS Inc.

Analysis of Abstraction and Structural Complexity Within Domain of Program Samples

¹Emília PIETRIKOVÁ (2nd year), ²Ivan HALUPKA (1st year)

Supervisor: ³Ján KOLLÁR

^{1,2,3}Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹emilia.pietrikova@tuke.sk, ²ivan.halupka@tuke.sk

Abstract—This article presents an approach for analysis of language abstraction and its complexity. Measurement results are presented, based on program samples, showcasing the effect of abstraction on language complexity, with contribution to a new approach in development of programming languages. All the presented measurements were performed within a specific domain of programs, providing sample derivation trees.

Keywords—Abstraction, list comprehension, programming languages, complexity, domain-specific languages.

I. INTRODUCTION

Programmers need to deal with a great amount of complexity [1]. With growth of software systems, expression complexity of their properties in a programming language mounts up as well. As the answer to complexity, higher levels of abstraction can be introduced. Abstraction allows expressing problems more simple by defining new, more abstract concepts that encapsulate complex expressions. This allows to hide implementation details. Therefore, a promising solution for growth of program complexity can be abstraction based on a language, allowing reduction of the complexity through definition of new, more abstract concepts and language constructions.

For example, expression $\frac{-b + \sqrt{b^2 - 4ac}}{2a}$ for computing one root of quadratic equation $ax^2 + bx + c = 0$ can be simplified by introducing new abstract concept — *discriminant* (D). This form can be even more simplified by defining abstraction corresponding to the whole expression (see Fig. 1).

This way of problem solution is divided into several levels, where each level provides abstractions for level above it (this can be called “stratified design” [2]). Provided, that lower levels are already in place, we can concentrate on problem solution and leave out implementation details. The role of abstraction within hierarchy of relations between hardware and software depicts Fig. 2.

Programming languages are also part of the hierarchy of abstraction levels. They provide a number of build-in abstractions that can be used to build programs. Moreover, they also provide ways to define new abstractions. For example, it is possible to define new functions, data structures and classes.

However, these standard ways of introducing new abstractions are often insufficient. In these cases, it is feasible to extend a language itself, and thus to use the method of metalinguistic abstraction [3].

The conception of programs as multiple levels of abstraction can be considered from a language perspective. This is a

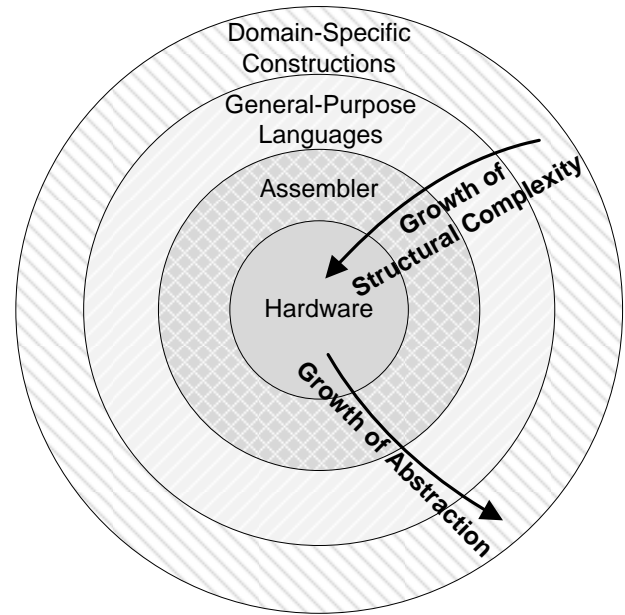


Fig. 2: The role of abstraction within hierarchy of software and hardware systems.

basic idea of Language-Oriented Programming [4], [5]. In this methodology, first step of program design is definition of high-level domain-specific language suitable for solving the problem. Next, this language is implemented and program itself is implemented using a new language. From this point of view, every level of abstraction is represented by a language, where each language is defined using a language on lower level.

Considering the importance of the concept of abstraction in programming, there are a lot of open questions remaining, particularly regarding automatic analysis and introduction of abstraction. In this article we are trying to find answers to these questions:

- Can effect of abstraction be measured?
- How can increasing of abstraction be automated?

To resolve both issues, analysis of sample programs is required. For this reason, a tool for program analysis from syntax perspective was developed. Section II presents an approach to measurement of abstraction effect based on syntactic analysis of programs.

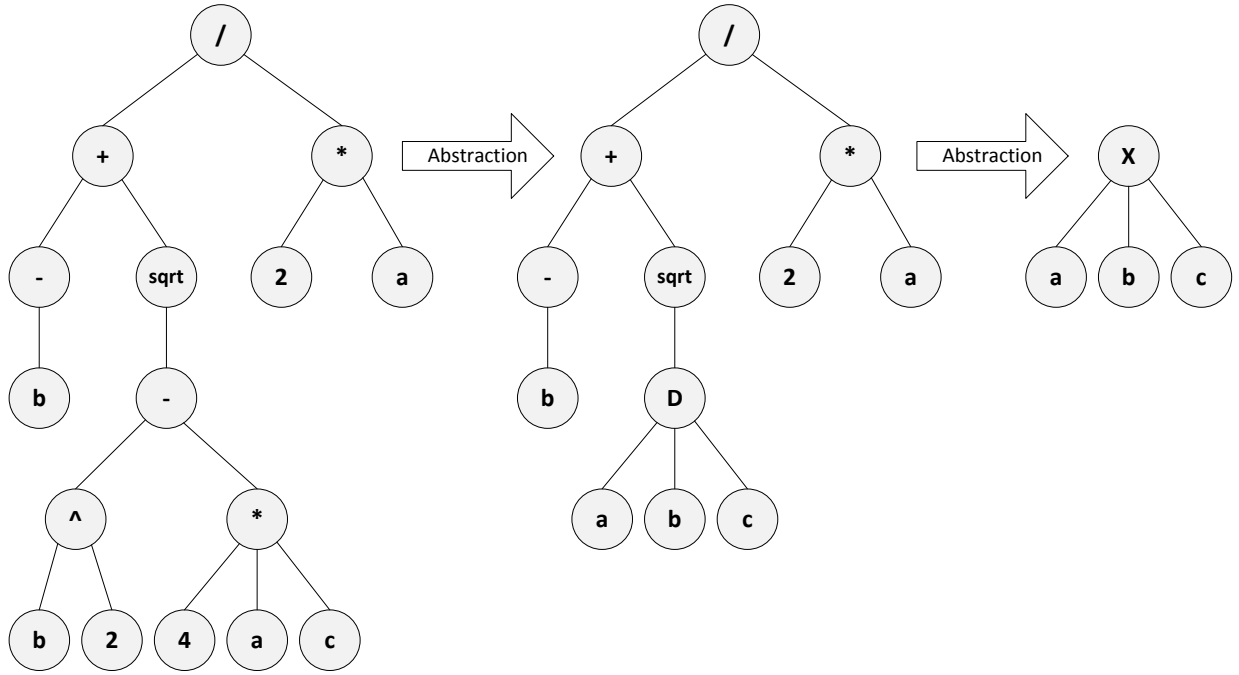


Fig. 1: Simplification of expression structure using abstraction.

II. MEASUREMENT OF ABSTRACTION

To measure the effect of abstraction, it is necessary to have an example of abstract construct. For the purpose of this article, list comprehension was chosen.

List comprehension (or set abstraction) is a powerful construct of the Haskell programming language that enables the use of notation equivalent to Zermelo-Fraenkel set notation.

List comprehension is a good example of abstraction, because it provides much more abstract notation for list manipulation compared to usage of other list manipulation operations. At the same time, every list comprehension expression can be translated into the form with lower level of abstraction.

A. List Comprehension

List comprehension LC can be defined as an expression $[E \mid Q_1, \dots, Q_n]$ with syntax presented in Fig. 3 [6]. If the E expression is of type T , list comprehension is then of type $[T]$.

LC	$::= [E \mid QS]$
QS	$::= Q_1, \dots, Q_n, \text{ for } n \geq 0$
Q	$::= G \mid F$
G	$::= p \leftarrow L$
p	$::= \text{pattern}$

Fig. 3: Syntax of list comprehension

List comprehension contains a list QS of qualifiers, separated by commas. Each qualifier can be either generator G or filter F . Filter F is a logical expression, and generator G produces patterns p of list L . If $p : T_p$, then $L : [T_p]$, where pattern p can be a variable or a constant of product type (e.g. tuple).

Translation of list comprehension is defined by translation scheme \mathcal{C} according to Fig. 4. Except for the lambda

abstraction, application of if operation is applied during implementation of the list comprehension, expressed as $\text{if } b \in_T e_F$. In the Haskell language, it is possible to represent it through expression of extended lambda language $\text{if } b \text{ then } e_T \text{ else } e_F$.

In the translation scheme, qs is a list of qualifiers and function h is as follows:

```

h          :: (a -> [b]) -> [a] -> [b]
h f []     = []
h f (x:xs) = f x ++ h f xs

```

It is possible to prove, that

```
h f = concat . (map f)
```

By expressing a simple list comprehension through extended lambda language according to the translation scheme, it is possible to practically ascertain the correctness of implementation on the basis of this scheme.

Theoretical analysis of the scheme accuracy based on concat and map functions might be simpler than analysis based on optimized function h . E.g. right side of equation (4) in the translation scheme:

```
h ( \p.C [ [ E | qs ] ] ) ( \mathcal{E} [ L ] )
```

can be expressed equivalently as:

```
concat (map ( \p.C [ [ E | qs ] ] ) ( \mathcal{E} [ L ] ))
```

because of the equation:

```
h f xs = (concat . (map f)) xs
```

and thus the following equation is true as well:

```
h f xs = concat (map f xs)
```

B. Experiment

Let us consider four list comprehension expressions:

- 1) $f = [x \mid x \leftarrow [1,2,3]]$
- 2) $f = [(x,y) \mid x \leftarrow [1,2,3], y \leftarrow [10,20,30]]$

$$\begin{array}{ll}
\mathcal{E} \llbracket [E \mid qs] \rrbracket & = \mathcal{C} \llbracket [E \mid qs] \rrbracket \\
\mathcal{C} \llbracket [E \mid] \rrbracket & = [\mathcal{E} \llbracket E \rrbracket] \quad (1) \\
\mathcal{C} \llbracket [E \mid F, qs] \rrbracket & = \text{if } (\mathcal{E} \llbracket F \rrbracket) (\mathcal{C} \llbracket [E \mid qs] \rrbracket) ([]) \quad (2) \\
\mathcal{C} \llbracket [E \mid F] \rrbracket & = \text{if } (\mathcal{E} \llbracket F \rrbracket) (\mathcal{C} \llbracket [E \mid] \rrbracket) ([]) \quad (3) \\
\mathcal{C} \llbracket [E \mid p \leftarrow L, qs] \rrbracket & = \text{h} (\lambda p. \mathcal{C} \llbracket [E \mid qs] \rrbracket) (\mathcal{E} \llbracket L \rrbracket) \quad (4) \\
\mathcal{C} \llbracket [E \mid p \leftarrow L] \rrbracket & = \text{h} (\lambda p. \mathcal{C} \llbracket [E \mid] \rrbracket) (\mathcal{E} \llbracket L \rrbracket) \quad (5)
\end{array}$$

Fig. 4: Translation scheme of list comprehension (set abstraction)

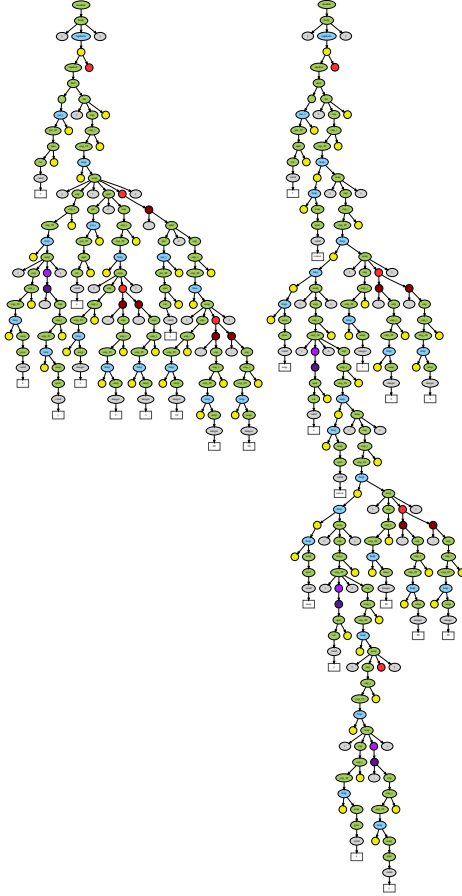


Fig. 5: Comparison of derivation trees of program 2

```

3) f = [ (x,y) | x<-[1,2,3], y<-[10,20,30],
        x>=2 ]
4) f = [ (x,y) | x<-[1,2,3], y<-[10,20,30],
        x>=2, y<=25 ]

```

We can translate them into less abstract form using translation scheme described in section II-A. For example, translated version of program 1 is as follows:

```
1 = concat (map (\x -> [x]) [1,2,3])
```

Translated version of program 4 is more complex:

```

f = concat (map (\x ->
  concat (map (\y ->
    if x>=2 then (if y <=25 then [(x,y)]
                      else [])
    else [])
    [10,20,30]))
  [1,2,3])

```

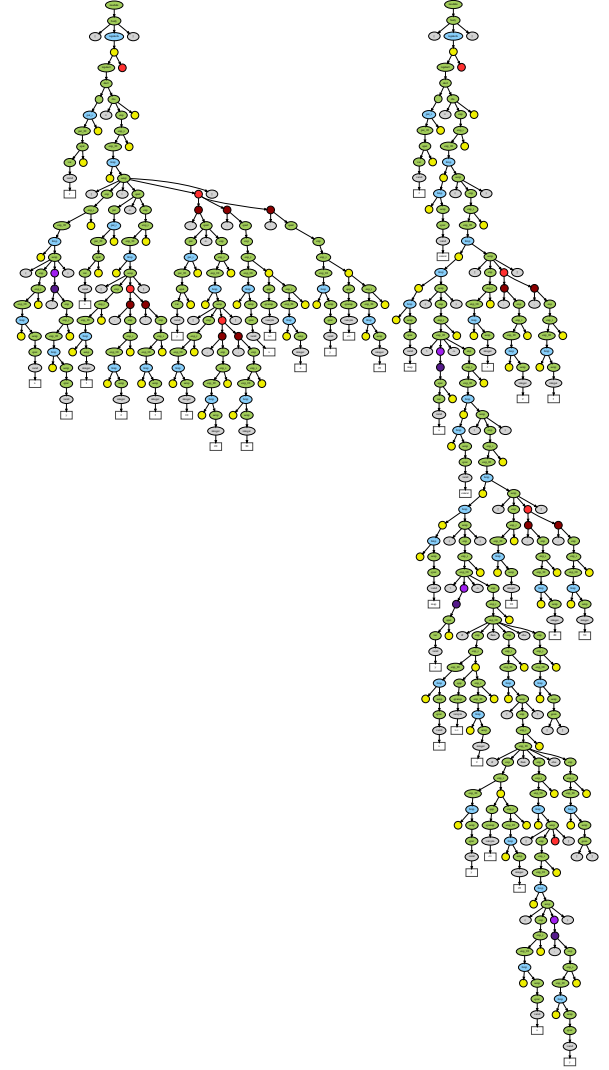


Fig. 6: Comparison of derivation trees of program 4

We can use these simple programs as an example of programs that express the same meaning using different levels of abstraction. The abstraction used in this case is language based abstraction — it is achieved by additional language construct (so called “syntactic sugar”).

Using the Haskell syntax analysis, we can produce derivation trees of these programs. Within Fig. 6, it is possible to compare derivation tree of list comprehension (program 4) to derivation tree of its translation (less abstract version).

Circles in the graph correspond to nonterminals and parts of complex nonterminals: alternative (green), repetition (red and purple), optional parts (yellow). Squares correspond to

terminals.

The more complex is the list abstraction, the wider is its derivation tree (more nodes). However, the tree itself remains fairly comprehensive.

On the other hand, each derivation tree representing less abstract expression of the same list gathers not only new nodes, but it is also compounded by transitions between nodes, thereby extending the tree into its length, and therefore reducing efficiency of the program result production.

C. Methodology of Measurement

If we want to analyze the effect of abstraction in programming language, it is necessary to define a method of measurement. Let us call our programs defining list comprehensions M_1 , M_2 , M_3 , and M_4 . We have detailed (less abstract) representations of these programs received using the translation scheme described in section II-A as well. Let us call them D_1 , D_2 , D_3 , and D_4 . For every program P , it is possible to create its derivation tree $T(P)$ based on language syntax.

Let $c(P)$ be the length of program code — the number of characters excluding white space. Then we can define ratio of abstraction of target (program code) as

$$Z_i = \frac{c(D_i)}{c(M_i)}$$

Let $n(G) = |V(G)|$ be the order of graph G — number of graph nodes. Then we can define ratio of abstraction of derivation as

$$T_i = \frac{n(T(D_i))}{n(T(M_i))}$$

If we analyze these ratios for a number of programs, we can see how abstraction impacts program length, size of derivation tree, and what the relation between these parameters is.

TABLE I: Results of the experiments

i	$c(M_i)$	$c(D_i)$	$n(M_i)$	$n(D_i)$	Z_i	T_i
1	16	29	88	113	1.81	1.28
2	34	60	159	208	1.76	1.31
3	39	76	181	244	1.95	1.35
4	45	95	203	289	2.11	1.42

Table I contains results of the measurement for presented programs. This simple experiment implies, that abstraction has greater impact on target form of the program (source code) than on its derivation, because relative change of length is greater in target form ($Z_i > T_i$).

This means that less effort on a more abstract production provides higher level of a target abstraction (in our case, target is the source form), what means that prevention of low levels of abstraction in the production can effectively increase the level of abstraction. Thus, higher level of abstraction increases the level of transparency and reliability of programs.

III. CONCLUSION AND FUTURE WORK

Analysis of the current state within application of functional programming languages proved that along with system development in various domains, there is a demand for the following language features [7], [8]:

- Increasing level of abstraction when expressing complex issues;
- Increasing expression ability of a language, and thus effectiveness of its application;
- Specialization of languages on specific domains of use;
- Increasing flexibility when using a language in other domains.

In this article we have shown that abstraction in programming languages has a great effect on programs. This effect has been analyzed on the process of source code derivation based on the Haskell 98 syntax. Experiment has been performed via Haskell syntax analysis, gathering needed information from Haskell programs and retrieving their derivation trees.

List comprehension and its translation corresponds directly to various levels of abstraction in the programs, and the provided derivation trees reflect that these levels of abstraction have strong impact on program derivation process. Analysis of target abstraction ratio and derivation abstraction ratio corresponds to conclusions proclaimed after producing and comparing derivation trees — that relative change of length is greater in target form than in the source form.

To make more significant conclusions, it is necessary to perform the experiment on greater set of programs. One helpful indication is that a slight variation of list comprehension within eight programs yields plausible results.

This supports the idea that is already known from the functional programming area regarding the fact that a language should be as simple as possible and, at the same time, it may be able to express a solution for any problem in a given problem area.

Further research will focus on methods of flexible language restructuring, based on a new form of language grammars. Then the aim should be resilient language adaptation on another application domain, what may have a contribution to construction or specialization of domain-specific languages as well.

ACKNOWLEDGMENT

The work presented in this paper was supported by SRDA Project of Slovak-Slovenian Research and Development No. SK-SI-0003-10 "Language Patterns in Domain-specific Languages Evolution".

REFERENCES

- [1] F. P. Brooks, "No silver bullet: Essence and accidents of software engineering," *IEEE Computer*, vol. 20, no. 4, pp. 10–19, april 1987.
- [2] H. Abelson and G. J. Sussman, "Lisp: A language for stratified design," Cambridge, MA, USA, Tech. Rep., 1987.
- [3] —, *Structure and Interpretation of Computer Programs*, 2nd ed., ser. MIT Electrical Engineering and Computer Science. The MIT Press, 1996.
- [4] M. P. Ward, "Language-oriented programming," *Software - Concepts and Tools*, vol. 15, no. 4, pp. 147–161, 1994.
- [5] S. Dmitriev. (2004, November) Language oriented programming: The next programming paradigm. Available: http://www.jetbrains.com/mps/docs/Language_Oriented_Programming.pdf.
- [6] J. Kollár, *Functional Programming (In Slovak)*, 1st ed. Elfa, 2009.
- [7] D. Astapov, "Using haskell with the support of business-critical information systems," *Practice of Functional Programming (in Russian)*, vol. 2, 2009.
- [8] A. Ott, "Using scheme in the development of "dozor-jet" family of products," *Practice of Functional Programming (in Russian)*, vol. 2, 2009.

Analysis of manufacturing systems using Petri nets

¹Matej ČOPÍK (2nd year), ²Ján ILKOVIČ (2nd year)
 Supervisor: ³Ján JADLOVSKÝ

^{1,2,3}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹matej.copik@tuke.sk, ²jan.ilkovic@tuke.sk, ³jan.jadlovsky@tuke.sk

Abstract — This article deals with the analysis of manufacturing systems that are modeled using Petri nets. The purpose of this article is present the tool for analysis of manufacturing systems in the design phase and helps to create real manufacturing systems. Article present an application that allows calculation of the reachability of wished state, the time transformation PN to the wished state and production costs on the basis of calculation for firing sequence. The existences of a firing sequence the feasibility of the selected state. With firing sequence is possible to calculate the time for transformation PN, performance of the manufacturing system, as well as production costs.

Keywords — Petri net, manufacturing system, incidence matrix, firing sequence

I. INTRODUCTION

A new manufacturing system is differing from the older and especially the complexity of the technology used. Today is trend to fully automate production to increase reliability, quality, productivity and efficiency. One of the main reasons for the automation of manufacturing systems is to reduce time to production and thereby increase production. When designing manufacturing systems, it is difficult to analyze the proposed project. Analysis of the project design phase often brings suggestions for improvement that is easily incorporated into the project. The analysis of the real systems brings better results, but improvement project is challenging. One of the aspects that can be analyzed is the time of the production cycle. If this is insufficient time already in the design phase, it is clear that the project not met demands for production. In this situation it is necessary to prevent early analysis and optimization of production systems.

In the first part of the article deals with the description of the model Flexible Manufacturing System (FMS) using Petri nets (PN).

The second part is devoted to the description of the application that analyzes the production system is modeled by PN. Application to the calculation of firing sequence analyzes production system.

II. DESCRIPTION OF THE MANUFACTURING PROCESS

A. Description of the model FMS

The production system can be considered a model of a fully automated production line, which was called FMS and serves as a learning model. The FMP is located at the Department of Cybernetics and Artificial Intelligence (KKUI), Technical University (TU) of Košice. This model is placed in the laboratory V147, which is at a Vysokoškolska street number 4. FMS is used in teaching and in the creation of bachelors, masters and doctoral theses. FMS is shown in Fig. 1 and a detailed description can be found in [1], [2] and [3].



Fig. 1. Flexible manufacturing system

B. Description of the model FMS using PN

The real model FMS is modeled using a generalized PN, which is shown on figure Fig.2. This model describes the activities of FMS. To monitor the time required to manufacture the product is necessary to extend this model for times of operation. The extension of the generalized PN with times we create T - timing PN. On model FMS is shown the basic calculations to be carried out to analyze the PN, which models the real manufacturing system.

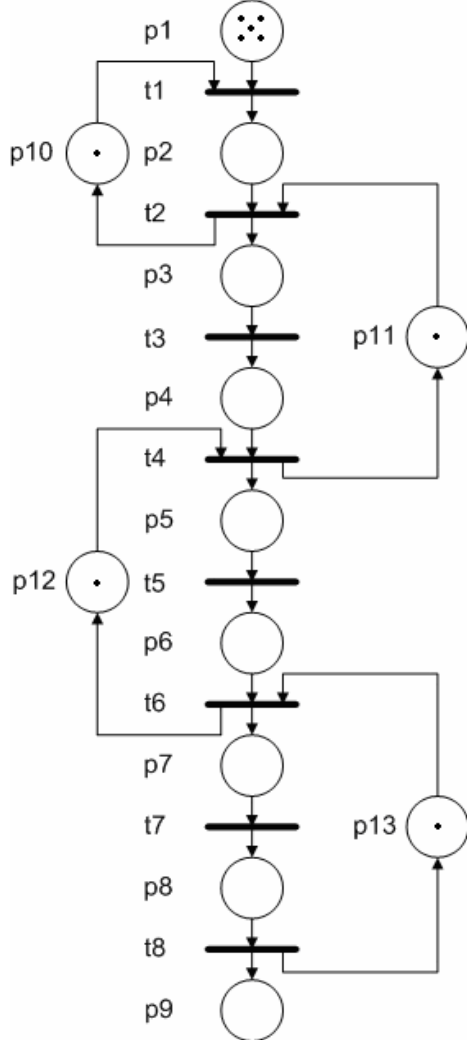


Fig. 2. PN modeling FMS

The places and transitions in the PN, these states and activities of FMS:

- p1 – a pallet is on input storage of post 3 and waiting for the release of the post 4,
- t1 - the movement of pallet on the post 4,
- p2 - a pallet of post 4 is ready to move to second conveyor and waiting for the release of the post 5,
- t2 - move pallet from first conveyor to second conveyor and then on post 5, release of post 4,
- p3 - a full pallet of post 5 is ready for emptying,
- t3 - emptying pallet on the post 5,
- p4 - empty pallet on the post 5 is ready to move and waiting for the release of the post 1,
- t4 - move pallet of post 1, release post 5,
- p5 – a pallet is on the post 1 and waiting for a folding shape,
- t5 - folding pictures in a pallet of post 1,
- p6 - a full pallet of post 1 is ready to move and waiting for the release of the post 2,
- t6 - transfer a pallet to post 2, release post 1,
- p7 - a pallet of post 2 is waiting for the end of control by camera,
- t7 - control by camera for a pallet of post 2,
- p8 - a pallet after control by camera on the post 2 is ready to move to the output storage on post 3,
- t8 - move pallet to the output storage on post 3, release post 2,
- p9 – a pallet is the output storage on post 3,
- p10 - post 4 is free,
- p11 - post 5 is free,
- p12 - a post is free,
- p13 - post 2 is free.

- p6 - a full pallet of post 1 is ready to move and waiting for the release of the post 2,
- t6 - transfer a pallet to post 2, release post 1,
- p7 - a pallet of post 2 is waiting for the end of control by camera,
- t7 - control by camera for a pallet of post 2,
- p8 - a pallet after control by camera on the post 2 is ready to move to the output storage on post 3,
- t8 - move pallet to the output storage on post 3, release post 2,
- p9 – a pallet is the output storage on post 3,
- p10 - post 4 is free,
- p11 - post 5 is free,
- p12 - a post is free,
- p13 - post 2 is free.

Model of manufacturing system must satisfy certain properties of PN such as reachability, liveness, conflictless, reversibility, and can not include own cycles.

PN of figure Fig. 2 can be described using matrices to describe the T - timed PN.

$$N = (P, T, pre, post, M_0, Tempo)$$

where $P = \{p1, p2, \dots, pm\}$ is a finite set of places,

$T = \{t1, t2, \dots, tn\}$ is a finite set of transitions,

pre - is direct incidence function, $pre: P \times T \rightarrow N$,

$post$ - is backward incidence function, $post: P \times T \rightarrow N$,

M_0 - is initial marking $M_0: P \rightarrow N$,

$Tempo(t_i) = d_i$ - specifies the time interval of the implementation of transitions, where $d_i \geq 0$.

The matrix pre a $post$, according to equation (1) calculate the incidence matrix C .

$$C = post - pre \quad (1)$$

$$C = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix} \quad (2)$$

III. ANALYSIS FMS USING PN

A. Insert the inputs to an application to the analysis PN

The application is a simple tool for the analysis of PN, which is described by matrices. Description by PN matrix is a simple procedure that can be done in the PN modeling programs as well as manually. Most modeling tools provide a description of the PN with matrices pre , $post$ and M_0 [5].

Matrix *pre* describes the multiplicity of input arcs into transitions and matrix *post* describes the multiplicity of output arcs from the transitions. Matrix M_0 describes the initial marking PN. These matrixes are input to an application to the analysis of PN. Matrix *pre* a *post* are usually sparse, so the manual entry of matrix application is sufficient to specify only non-zero values. Example of insert the matrixes into the application are shown on figure Fig. 3.

Fig. 3. Flexible manufacturing system

Matrixes after inserting into the application are stored and calculated the incidence matrix C for the PN. Reviews of matrixes are shown in a separate window that is on Fig. 4.

Fig. 4. Reviews of inserting matrix

The matrixes after the inserting can be edited. If the matrixes are correctly insert, can by go to the analysis of PN.

B. Analysis PN using application

The first step should be to verify the existence of the desired state of PN, but it is not a prerequisite.

For any state of PN is possible to calculate its reachability. Reachability state of PN can be detected from existence the firing sequence. If there existing firing sequence, which transforms the PN from the initial to the desired state then desired state is reachable. Firing sequence can take only nonnegative values of whole numbers. Firing sequence can be calculated using the modified state equation (3).

$$C^{-1} * (M_x - M_0) = \sigma \quad (3)$$

where M_x – is desired state,

M_0 – is initial marking of PN,

C – is incident matrix,

σ – is firing sequence.

The application after inserting the desired state is calculated the firing sequence, as is shown on Fig. 5. If the desired state is not reachable, the application displays an error message.

Fig. 5. Calculating firing sequence

For reachable states of the PN is possible make further analysis. The application allows to analyze the T and P timed PN. The PN model can specify time intervals for different places or transitions. According the time intervals are calculated time of transition for desire state of PN. The application calculates the costs necessary for the transformation of PN in the desired state. However, it must specify costs for each places or transitions of PN. On Fig. 6. is shown the screen, where is possible analyze the time and cost needed to transform the PN. Analysis of different states PN is impossible to compare the calculated values. Analysis of values is possible to assess whether the results obtained are sufficient or not [4].

Fig. 4. Screen for analysis production time and cost

The calculated value is the value of the real modeled manufacturing system. If reachable results are sufficient, it is possible to move to the implementation of the production system. However, if the results obtained are not sufficient, we need to make changes in design and analysis execution again.

IV. CONCLUSION

The area of the design and implementation of production systems is nowadays becoming increasingly important. Major role in this respect have new factories of the future, which seek to minimize human intervention in the production process. In the design of production systems, increasingly thorough analysis is gaining prominence. One of the methods of analysis of production systems is the use of PN. Using the model of the production system, one can analyze some properties such as reachability state of the manufacturing process, the production time or production costs. Presented application is a simple tool for the analysis of PN, which modeled real manufacturing systems. Based on the results of the analysis of PN, one can decide on the implementation of the manufacturing system or its re-projection.

ACKNOWLEDGMENT

This contribution is the result of the Scientific Grant Agency of the Slovak Republic and of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020), supported by the Research and Development Operational Program funded by the ERDF.

This work has been supported by the Scientific Grant Agency of the Slovak Republic under project Vega No.1/0286/11 Dynamic Hybrid Architectures of the Multiagent Network Control Systems.

REFERENCES

- [1] ČOPÍK, M. 2010. Návrh a realizácia riadenia pružného výrobného systému: Diplomová práca (supervisor: Ján Jadlovský). Košice: TU, 2010. 75 s.
- [2] ČOPÍK, M. – ILKOVIČ, J. 2011. Proposal control of manipulators controlled via technological network DeviceNet In SCYR 2011 : 11th

Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice [CD-ROM]. Košice : Technická univerzita, 2011. s. 335-338. ISBN 978-80-553-0644-5

- [3] ČOPÍK, M. – LACIŇÁK, S. – JADLOVSKÝ, J. 2010. Návrh a realizácia riadenia pružnej výrobnéj linky. In Electrical Engineering and Informatics: Proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice [CD-ROM]. Košice : Technická univerzita, 2010 s. 561-564. ISBN 978-80-553-0460-1.
- [4] ERTEL, D. – DROZD, P. 2007. Petriho siete a ich využitie pri analýze správaní sa modelu výrobnéj linky: ASR ŠVOČ. Bratislava: STU, 2007. 6 s.
- [5] KORDIC, V. 2008. Petri net theory and applications. Vienna: I-Tech Education and Publishing, 2008. 544 s. ISBN 978-3-902613-12-7

Anomalies detection in user's motion patterns in a smart-home environment

¹Marek Novák (3rd year), ²Martin Kapa (3rd year)
 Supervisor: František Jakab

^{1,2}Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹marek.novak@tuke.sk, ²martin.kapa@tuke.sk

Abstract—The paper presents a model for anomaly behaviour detection service. The service detects unusually long periods of inactivity, lacking activity, unusual presence and changes in daily rhythm. Service is working in a proactive way, meaning it can learn user's habits without statically defined rules. The model is based on an unobtrusive approach - ambient presence sensors. For anomaly detection we have employed unsupervised clustering technique Self Organizing Maps and next activity prediction based on Markov model. Finally we present a short experimental study realized on a dataset provided by MavHome project.

Keywords—anomaly detection, behavioural patterns, classification, Markov model, Self Organizing Maps, smart-home

I. INTRODUCTION

One of the components provided by MonAMI (*Mainstreaming on AMbient Intelligence*) system, in which we participated on development, is a service called *ZoneSURE* [1]. The aim of the service is to provide a notification to a carer if presence of the user in a predefined time interval and place is not recognized when expected, or is detected when should be not. The carer is immediately informed of such a situation. Despite the fact that time intervals and locations (usually rooms) were defined for the service statically without any flexibility and adaptability, the users evaluated this service very positively after trials taking 3 months in 20 households in Košice, Slovakia [2].

This paper presents an approach to a similar service with pro-active functionality, where behaviour pattern of a user is learned by the service, instead of statical configuration based on inquiries or personal interview.

The data used in our work are from unobtrusive sensors as passive infra-red sensors (PIR) or bed/sofa pressure sensors. We argue, based partly on a feedback of our previous work [2], that ambient sensing is far more accepted by users than different types of tags on a body or devices like cameras.

Resulting service is not aimed for saving lives in highly critical situations, where the notification is a must in a few minutes. Rather we want to provide alerts indicating to a carer an unusual situation forcing him to undertake needed steps. Nevertheless, this could save life, when for instance elderly living alone lost consciousness, or is unable to move and unusually long inactivity is detected.

Section II. presents similar works and discusses the difference between obtrusive and unobtrusive approaches. Section III. presents our model based on SOM classification algorithm and simple Markov prediction model. In section IV. we presents a preliminary study on a dataset from MavHome project [3].

II. RELATED WORK

Research in supporting independent living of elderly people is highly actual. Especially the area of *activities of daily living* (ADL) is very dynamic nowadays [4]. Various projects are devoted to monitoring and collecting information by means of sensors [5], activity classification [6] or proactive reservation of resources [7].

A. Obtrusive and Unobtrusive Approach

Techniques used in ADL may differ in various aspects. Most visible is a type of sensors, or in other words, mean of data collection. Unobtrusive sensors are located in the living environment as PIR sensors, motion detectors, door/window entry point sensors, electric usage power sensors, pressure sensors, etc. The user is not disturbed with the technology and after a short time gets absolutely accustomed to it.

On the other hand, there are obtrusive techniques based on accelerometers or tags attached to a user body [8] and rich sensors as cameras or microphones [9]. Obtrusive sensors highly interfere common living habits, where user is either required not to forget attach tags on a body each day or he faces a loss of privacy. Although obtrusive sensors provide more informative output, the aforementioned drawbacks decreases final benefit of the service, since the user is discouraged to use it.

B. Lower or Higher Data Analysis

Another distinction is based on a processing of the gathered data. Some approaches use directly sensory data to find outliers as seen in a work of J. Weisenberg et al. [10]. Inputs to a learning algorithm in their work are time stamps of sensors' triggeres and intervals of inactivity. The model does not have any knowledge if the user is sitting, standing, walking, etc. It recognizes only whether there is a motion or not.

B. Kaluža and M. Gams [8] uses an approach of transforming behaviour traces into a spatial-activity matrix (Fig. 1) resulting in a graphical representation of daily living dynamics. The matrix represents a function of mapping higher level activities as sitting, standing in a particular room to a color. The warmer the color, the higher frequency of activity was detected. The outliers are determined by differences in activity values (color) in different days. This approach belongs to a category of higher level activities detectors, since the sensory data needs to be firstly classified into activities, which are consequently used for anomaly detection.

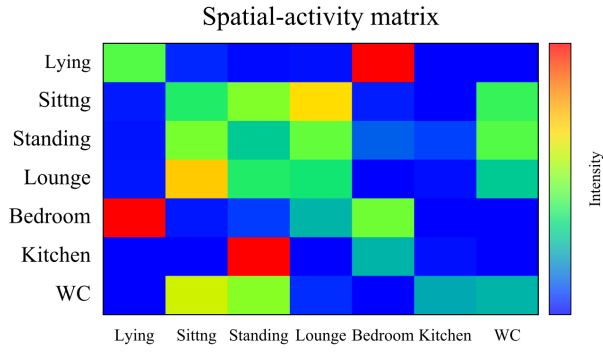


Fig. 1. Visualization of the spatial-activity matrix mapping of rooms and activity to a color. Warmer colours represents higher frequencies of activity [8].

C. Our contribution

We would like to avoid the usage of obtrusive sensing, because from the results of MonAMI project it is evident that users do not want to use neither hardware attached to their body, even if it is small in a shape as panic button (it has taken a long time to persuade the beneficiaries to wear it), nor the cameras, because they are afraid of privacy loss. The higher activity classification is therefore impossible.

Similar works oriented on raw data analysis, to our best knowledge, currently covers only some specific anomalies as unusual long inactivity.

Our novelty is in movement classification into activities, however without the assumption of a higher level meaning (as sitting, standing, etc.) connected with next activity prediction based on Markov model.

III. ANOMALY DETECTION OF BEHAVIOURAL PATTERNS

We aim to design an unobtrusive outlier detection service based on motion (presence) data classification and next class prediction. In the following work we will consider only data from presence sensors, however also others may be added in future (e.g. electric device usage). In addition, there is a presumption a person lives alone and pets are not considered, though this could be solved with a special RFID tag on a pet's collar.

In Fig. 2 there is a snapshot of synthetic daily activity of a person in two rooms. From the picture one can easily identify the period of sleeping, having a nap in the afternoon and making morning and night hygiene. Obviously, the real data will exhibit more noise.

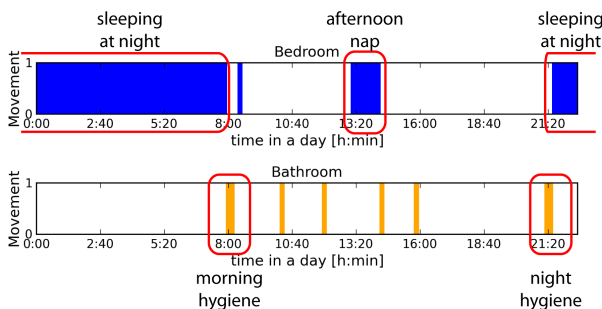


Fig. 2. A sample of synthetic daily living activity for a single day

A. Anomaly Classification

For classification we will use data comprising of three values, similar to the work of A. Lotfi et al. [11]

- 1) *sensor location* (l) - the room, where the sensor is installed
- 2) *start time* (s) - time when presence in the room was detected
- 3) *duration* (d) - the length of time interval, while person stays in this room (active or inactive)

For each room there is created a separate function of mappings start times to durations. The values of these functions are consequently classified by clustering algorithm which results in a unique classification for each room. The sample of such a classification is depicted in Fig. 3 filled with synthetic data for one month. The graph shows that person is used to go sleep with a mean at 22:20 and sleeps around 8 hours. The person is also having a nap with a mean at 13:00 and duration about 1 hour. There are located three anomalies indicating the person waken up after 4 hours.

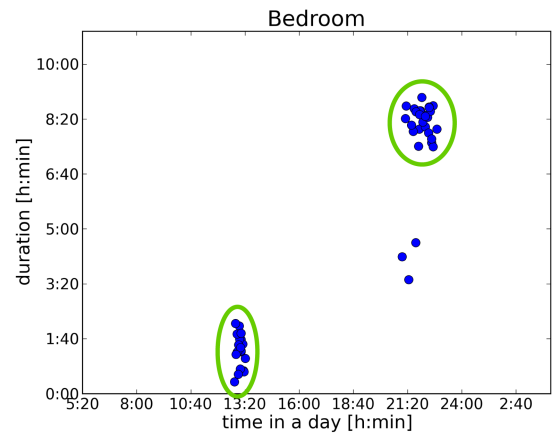


Fig. 3. A sample of synthetic data of one month in a bedroom

A start time value is given by the time a person enters a room. Duration is given by the length of a time interval when person is located in a room either actively or inactive, so the interval is ended when transiting from one room to another.

B. Types of Anomalies to be Detected

Following are types of anomalies with example situations, that system deals with:

- *Unusual long inactivity* - This could happen, when a user falls or loses consciousness. Duration for a specific time will be unusually long, therefore evaluated as anomaly.
- *Not present when should be* - When user is expected to be at home, but still doesn't return.
- *Unusual presence* - The example situation is, when user is sitting in a living room at night, but is expected to be sleeping. It may indicate some health problem.
- *Unusually short activity* - Short activities like transitions between rooms are not included in classification, therefore not considered in the model.
- *Changes in day rhythm* - The user used to prepare breakfast after waking up. When sitting long time in a living room without breakfast could indicate e.g. digestion problem. To find anomalies of this type a simple Markov model for next activity prediction is employed.

C. Self Organizing Maps

Classification may be done by various clustering techniques as: *K-means*, *Fuzzy c-means*, *Neural Gas*, etc. We have employed a technique *Self Organizing Maps (SOM)* for unsupervised learning, because of the fact, it does not require a number of clusters to be defined in advance as it is with the *K-means*. The number of clusters for each room is different and could be hardly guessed in advance.

SOM is a special type of artificial neural network, usually used for a projection of given data items onto two dimensional grid, firstly developed for the visualization of distributions of metric vectors, but adaptable to clustering of any data items, for which mutual distance is measurable [12].

Data item input in our case is 2-dimensional:

$$x(t) = [a(t), d(t)] \quad (1)$$

where $a(t)$ is time when entered a room, $d(t)$ is duration of staying in that room and t is the index in the input sequence. Map's node is defined as m_i which is a weight vector in artificial neural network (ANN). In a smoothing-process new value for i^{th} map's node is:

$$m_i(t+1) = m_i(t) + \alpha(t)h_{ci}(t)[x(t) - m_i(t)] \quad (2)$$

Here the index t is both an iteration in computing a model as well as input sequence number. $\alpha(t)$ is a scalar factor decreasing with t . The factor $h_{ci}(t)$ is a neighborhood function of models $m_i(t)$ depending on a distance between $x(t)$ and the winner of a particular sample, deeper described in [12]. Algorithm for SOM is sketched in Alg. 1

Algorithm 1 Schema of a SOM algorithm

```

 $m_1$  = random weight vector
 $t = 1$ 
while  $t < limit$  do
  for each node  $m_i$  in a map do
     $d$  = euclidean distance( $m_i(t)$ ,  $x(t)$ )
    track  $m_i(t)$  with smallest  $d$ 
     $m_i(t+1) = m_i(t) + \alpha(t)h_{ci}(t)[x(t) - m_i(t)]$  (2)
  end for
   $t = t + 1$ 
end while
 $\triangleright d(m_i(t), x(t))$  is euclidean distance to measure
the similarity between input vector and a node counted as
 $d(r, x(t)) = \sqrt{(m_{i1} - x_1(t))^2 + (m_{i2} - x_2(t))^2}$ 

```

D. Next Activity Prediction

Activity prediction in our model is based on an observable first-order Markov model (3), where S is the set of states and A is transition probability matrix [13].

$$\begin{aligned} M &= (S, A) : \\ S &= \{q_1, \dots, q_n\} \\ A &= (p_{ij}) : 1 \leq i, j \leq n \end{aligned} \quad (3)$$

Markov model is a stochastic model assuming Markov property (4), where any future state of the process depends only upon the present state, not considering previous history.

$$p_{ij} = P(q_t = S_j | q_{t-1} = S_i) \quad (4)$$

The states of a model are given by activities recognized by clustering algorithm for all rooms. The activity is therefore specified by room, time and duration. Markov model does not utilize any of these values, we mention it just for better understanding of transitions between states in Markov model as transitions between different rooms in different times.

The probability of transition from activity a_i to a_j is counted as fraction of sum of transitions between these two states and all transitions from activity a_i (5).

$$p_{ij} = \frac{\sum(a_i \rightarrow a_j)}{\sum(a_i)} \quad (5)$$

The model evaluates as an outlier transitions which are highly improbable (with a very small p_{ij}).

IV. EXPERIMENTAL STUDY

A. Dataset from MavHome Project

To create a consistent and suitable ADL dataset is a difficult task requiring reliable sensor network installation and a long time of data acquisition. Similar works utilize databases created in *living labs*, currently evolving throughout the Europe [14]. Living lab is a user centric methodology composed of a laboratory (could be flat or house) with sensors and actuators installation networks, where the behaviour and interaction of people is studied by several research areas.

B. Dataset from MavHome Project

Currently not having our own suitable dataset of an elderly person, we decided to use a dataset created within MavHome project [3]. They have monitored the behaviour of a student living alone in a two rooms apartment plus kitchen, toilet and bathroom. Obviously, the behavioural pattern of a young student will not show such regularities as it is expected for an elderly. Nevertheless, for the presentation of our ideas it is sufficient.

The data in MavHome projects were collected in a two months period in an apartment equipped with 36 motion sensors and 6 reed switch. Dense distribution of the sensors in a flat area is depicted in Fig. 4. Since we want to monitor just presence and transitions between different areas (rooms) of a flat, we joined sensors in one room to one logical sensor.

C. Results

For training a SOM classification we have used the first month of MavHome dataset. Testing was done on the same dataset with artificially added anomaly behaviour. System was able to find anomaly in more than 80% cases. Naturally, we are aware that for system verification more testing is required in more real conditions. This we would like to focus on in our future work.

Daily activity of a student from MavHome dataset measured in one month interval is depicted in Fig. 5. As one can see, the living habits of a student are rather flexible, but the graph shows the data before clearing shorts transitions. After clearing, only wider parts with higher opacity will remain (indicated by gray ellipses). We assume that behaviour of an elderly is more stationary than that of a student.

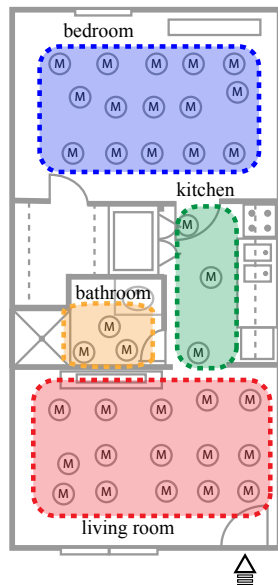


Fig. 4. Movement sensors' placement in a flat area in MavHome project

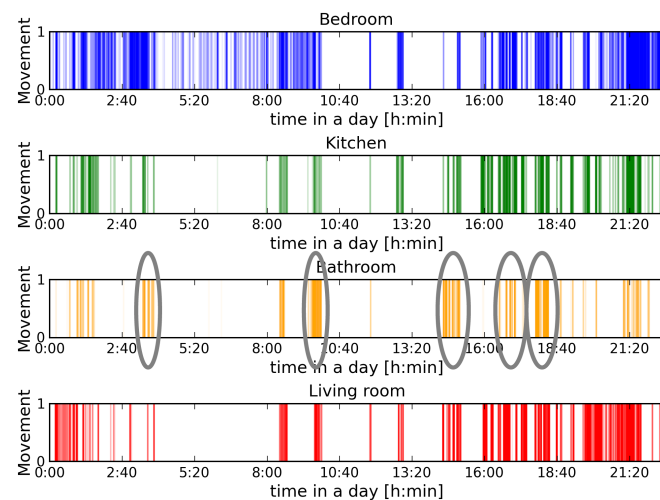


Fig. 5. Activity of a student from MavHome dataset before clearing. Gray ellipses indicate selected activities.

V. CONCLUSION

In this paper we have presented a model for finding anomalies in user's motions patterns. Input data to SOM classification are start time of an activity and duration of an activity either in active or inactive way (user is moving or he just stays inactively in a room). Duration is given by transitions among rooms. With this approach we are able to classify unusual long periods of inactivity and unusual presence. First level Markov model is used to find anomalies in user's daily rhythms.

We have realized a preliminary experimental study on a dataset created within MavHome project. However, the dataset monitors a young student with more noisy behaviour as it is anticipated for elderly. The more exhaustive experiments are needed, which we will work in the future.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] D. Šimšík, A. Galajdová, D. Šiman, M. Novák, and P. Galajda, "Services for seniors - experience of testing in slovakia field trials," *Assistive Technology Research Series: Everyday Technology for Independence and Care - AAATE 2011*, vol. 29, pp. 1082–1089, 2011.
- [2] A. Galajdová, D. Šimšík, M. Andrášová, J. Bujňák, and S. Krajňák, "Socio-economical aspects of ict based services for seniors in slovakia," *Assistive Technology Research Series: Everyday Technology for Independence and Care - AAATE 2011*, vol. 29, pp. 286–293, 2011.
- [3] G. Youngblood and D. Cook, "Data mining for hierarchical model creation," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, no. 4, pp. 561–572, 2007.
- [4] N. Noury, J. Poujaud, A. Fleury, R. Nocua, T. Haddidi, and P. Rumeau, "Smart sweet home... a pervasive environment for sensing our daily activity?" *Activity Recognition in Pervasive Intelligent Environments*, vol. 4, pp. 187–208, 2011.
- [5] H. Medjahed, D. Istrate, J. Boudy, and B. Dorizzi, "A fuzzy logic system for home elderly people monitoring (emutem)," in *Proceedings of the 10th WSEAS international conference on Fuzzy systems*. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2009, pp. 69–75.
- [6] D. S. Wonjoon Kang, Dongkyoo Shin, "Detecting and predicting of abnormal behavior using hierarchical markov model in smart home network," in *Industrial Engineering and Engineering Management (IEEM)*, 2010, pp. 69–75.
- [7] T. Miklušček and M. Gregor, "Person movement prediction using artificial neural networks with dynamic training on a fixed-size training data set," *Applied computer science : management of production processes*, vol. 7, no. 2.
- [8] B. Kaluža and M. Gams, "An approach to analysis of daily living dynamics," *World Congress on Engineering and Computer Science, WCECS 2010*, vol. 1, pp. 485–490, 2010.
- [9] T. Choudhury, M. Philipose, D. Wyatt, and J. Lester, "Towards activity databases: Using sensors and statistical models to summarize people's lives," *IEEE Data Eng. Bull.*, 2006.
- [10] J. Weisenberg, P. Cuddihy, and V. Rajiv, "Augmenting motion sensing to improve detection of periods of unusual inactivity," in *Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*, ser. HealthNet '08, 2008, pp. 2:1–2:6. [Online]. Available: <http://doi.acm.org/10.1145/1515747.1515751>
- [11] A. Lotfi, L. Caroline, S. Mahmoud, and M. Akhlaghinia, "Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14.
- [12] T. Kohonen and T. Honkela, "Kohonen network," 2007, accessed: March 2012. [Online]. Available: http://www.scholarpedia.org/article/Kohonen_network
- [13] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, 1989, pp. 257–286.
- [14] "European Network of Living Labs - knowledge Center," accessed: March 2012. [Online]. Available: <http://knowledgecentre.openlivinglabs.eu/>

Audio Content Classification using SVM Binary Decision Trees

Jozef VAVREK (2nd year)

Supervisor: Anton ČIŽMÁR

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

jozef.vavrek@tuke.sk

Abstract—Audio content classification is an important issue in current audio processing and can be implemented in many audio applications, such as indexing and retrieving. In this paper we present broadcast news audio classification algorithm, based on support vector machine - binary tree (SVM-BT) architecture. Five audio classes are considered: pure speech, speech with music, speech with environment sound, pure music and environment sound. The experimental results show the efficiency of used SVM-BT architecture in broadcast news audio classification task, especially for speech nonspeech discrimination.

Keywords—Classification, segmentation, feature extraction, support vector machines binary decision trees.

I. INTRODUCTION

A big amount of audiovisual data has been archived in broadcast news (BN) databases. An effective segmentation, classification and indexing algorithm may facilitate the retrieval of desired audio or video segments. This paper is focused on audio stream classification via Support Vector Machine binary decision tree (SVM-BT) architecture that ensures audio content processing in terms of discrimination various types of acoustic events. It is used for processing and retrieving of broadcast news data, a mass media with rich audio content. Audio content of broadcast news data contains not only single type of acoustic event (pure speech, music), but also mixed sounds (speech and music with background). That audio stream consists of the following audio events:

- *pure speech* - single anchor reports in studio, field speech in a quiet environment,
- *speech with environment sound* - field speech or interviews and telephone conversations in noisy environment,
- *speech with music background* - jingles at beginning and at the end of news with anchor speech, commercials with speech,
- *pure music* - jingles, music in commercials,
- *environment sound* - all types of acoustic events in BN audio stream, e.g. sound from machines, birds, water, wind, crowds, etc.,
- *silence* - silent intervals between different speakers and jingles before speech.

The main effort in content-based audio analysis tends to separation of two types acoustic events, namely speech and music (non-speech). A human listener can easily recognise various types of speech and music by listening only a small parts of audio signals, but it is a very difficult problem for computer processing.

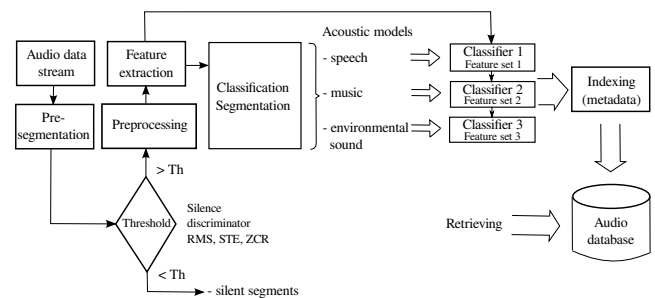


Fig. 1. Basic concept of automated classification and segmentation system for broadcast news audio data

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 of audio signal by 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: Section II deals with segmentation and feature extraction. Section III provides description of the used SVM binary tree topology and section IV discuss realized experiments and obtained results. Finally, the section V gives our conclusions and shows future directions.

II. SEGMENTATION AND FEATURE EXTRACTION

Audio segmentation is in general the task of dividing 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 contains the same type of acoustic signal (speech, music or environmental sound). Therefore segmentation ability strongly depends on the segment length of processed audio stream.

A. Pre-segmentation

Segmentation is often connected with the task of preprocessing or pre-segmentation of audio signal. In that meaning the segmentation is taken as determination of segments with equal length, usually by rectangular window. The rule of homogeneity is not always guaranteed. Time domain statistical parameters such as mean, variance, Root Mean Square (RMS) of Short-time Energy and also Zero-Crossing Rate (ZCR) are computed within these segments. Typical length of one segment is 1s. Pre-segmentation is followed by frame-based segmentation, where each segment is further divided into overlapped frames, using Hamming window, in order to avoid spectral distortions.

B. Feature Extraction

The primary and important task of the BN segmentation and subsequent classification is to extract features that objectively represent the audio content. All possible and often used features can be categorized into four classes [8]:

- *Time domain 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, zero-crossing rate (ZCR),
- *Frequency domain / Spectral features* - most of them are defined in MPEG-7 standard [9], for example spectral roll-off (ROF), audio spectrum centroid (ASC), spectral flux (SF), spectral kurtosis (SK), audio spectrum spread (ASS) and flatness (ASF), etc.,
- *Cepstral features* - are derived from human perceptual (auditory) system, such as Mel Frequency cepstral coefficients (MFCCs), 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 (DP), spectral peak duration (SPD), pitch tuning (PT), average pitch-density (APD) and relative tonal power density (RTPD).

Following subsections describe the basic descriptors and feature extraction methods used in our experiments. These descriptors were chosen according to the sufficient results of classification accuracy of audio data that were presented in [8], [10].

1) *Short-Time Energy*: STE of a discrete signal is defined as the sum of squares of the signal samples $x[n]$ normalized by the segment length N , often described as frame.

$$STE = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n]. \quad (1)$$

2) *Zero-Crossing Rate*: ZCR of a frame is defined as the number of times the audio waveform changes from positive to negative within the duration of the frame (sign-changes).

$$ZCR = \frac{1}{2} \sum_{n=0}^{N-1} |sgn(x[n]) - sgn(x[n-1])|. \quad (2)$$

3) *Audio Spectrum Centroid*: ASC [11] is computed from the modified power coefficients and gives the information about the shape of the power spectrum. In the Eq. (3) $P'(k')$ represents the power spectrum and $f'(k')$ represent corresponding frequencies.

$$ASC = \frac{\sum_{k'=0}^{(N_{FT}/2)-K_{low}} \log_2\left(\frac{f'(k')}{1000}\right) P'(k')}{\sum_{k'=0}^{(N_{FT}/2)-K_{low}} P'(k')}. \quad (3)$$

ASC descriptor indicates whether in a power spectrum are dominated low or high frequencies and can be regarded as an indicator of the 'brightness' or 'sharpness' of the audio signal.

4) *Audio Spectrum Spread*: ASS [11] describes concentration of the spectrum around the centroid and is defined as:

$$ASS = \frac{\sum_{k'=0}^{(N_{FT}/2)-K_{low}} \left[\log_2\left(\frac{f'(k')}{1000}\right) - ASC \right]^2 P'(k')}{\sum_{k'=0}^{(N_{FT}/2)-K_{low}} P'(k')}. \quad (4)$$

Lower spread values means that the spectrum is concentrated near the centroid, whereas a high value reflects a distribution of a power across a wider range of both sides of the centroid.

5) *Audio Spectrum Flatness*: ASF [11] can be defined as the ratio of the geometric and the arithmetic means. of the power spectrum coefficients within specific frequency band b :

$$ASF(b) = \frac{\sqrt[hiK'_b - loK'_b + 1]{\prod_{k'=loK'_b}^{hiK'_b} P_g(k')}}{\frac{1}{hiK'_b - loK'_b + 1} \sum_{k'=loK'_b}^{hiK'_b} P_g(k')}, \quad (1 \leq b \leq B). \quad (5)$$

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.

6) *Spectral Roll-off*: ROF [11] point tells how much of the frequencies are concentrated below 85% of the window's energy K_{roll} . It is often used as an indicator of the skew of frequencies in a window:

$$\sum_{k=1}^{K_{roll}} |S(k)| = 0,85 \sum_{k=1}^K |S(k)|, \quad (6)$$

where $S(k)$ represents spectrum of the signal at particular frequency bin k . RF is used to distinguish voice speech from unvoiced music, which has a higher roll-off point because their power is better distributed over the subband range.

7) *Mel-Frequency Cepstral coefficients*: MFCCs [12] are a compact representation of the spectrum of an audio signal taking into account the nonlinear human perceptual system, as described by the mel scale:

$$Mel(f) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right), \quad (7)$$

where f is frequency in Hertz. They are one of the most used features in speech recognition and have recently been proposed to analyze and represent musical signals. MFCCs are computed by grouping the Discrete Fast Fourier Transform (DFFT) points of each frame into a set of 13 coefficients, using bank of triangular shaped filters. Then the logarithm of the coefficients is taken, and the Discrete Cosine Transform (DCT) is applied to decorrelate them.

III. SVM BINARY TREE ARCHITECTURE

SVMs [13] have in recent years drawn much interest for audio classification due to their generalization ability and superior performance in various pattern classification tasks. 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 the SVM technique. In this sense, SVM represents specific type of discriminative function that do not model the whole distribution, but the most discriminative regions of the distribution, e.g., class boundary or margin. Discriminative function is modeled by linear separating hyperplane with maximal or soft margin. If the training data from different classes cannot be linearly separated in the original input space, the SVM at first nonlinearly transforms the original input space into the high-dimensional *feature space*. This transformation can be achieved by using various nonlinear mappings such as: polynomial, sigmoid as in multilayer perceptrons, RBF mappings having as basic function radially symmetric function, i.e. Gaussian and spline functions. The resulting hyperplane in feature space will be optimal in the sense of being maximal margin discriminative function with respect to training data.

Training data are represented in the form of N -dimensional vectors:

$$(x_1, y_1), \dots, (x_l, y_l) \in X \times \{\pm 1\}, \quad (8)$$

where X is some nonempty set of patterns x_i (sometimes called cases, inputs, instances, samples) and labels y_i . Among all the hyperplanes that minimize the training error, classifier have to find the one with the largest margin in the form:

$$d(\mathbf{w}, \mathbf{x}, b) = \langle \mathbf{w} \cdot \mathbf{x} \rangle + b = \sum_{i=1}^l w_i x_i + b, \quad (9)$$

where l represents all the training data, $\mathbf{w}, \mathbf{x} \in R^n$ and the scalar b is called *bias*. After successful training stage, using obtained weights, the learning machine produces output o , according to an *indication function*, given as:

$$i_F = o = \text{sign}(d(\mathbf{w}, \mathbf{x}, b)), \quad (10)$$

where o is the standard notation for the output from a learning machine.

One basic idea in designing nonlinear SVMs is to map input vectors $\mathbf{x} \in R^n$ into the high-dimensional feature vectors by *kernel functions* $K(\mathbf{x}_i, \mathbf{x}_j)$. The mostly used kernel functions are shown in the Tab. I.

Basic concept of the binary tree architecture is depicted in the Fig. 2. Each node represents one binary SVM classifier that realizes separation of two classes +1 and -1. It follows that the multi-class classification needs to train maximally $(K - 1)$ SVMs for K -class problem. Therefore, it is more efficient in

TABLE I
KERNEL FUNCTIONS

Kernel function	Label
$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle$	Linear
$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle + 1)^d$	Polynomial of degree d
$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \ \mathbf{x}_i - \mathbf{x}_j\ ^2)$	Gaussian radial basis function
$K(\mathbf{x}_i, \mathbf{x}_j) = (\rho \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle + r)$	Sigmoid

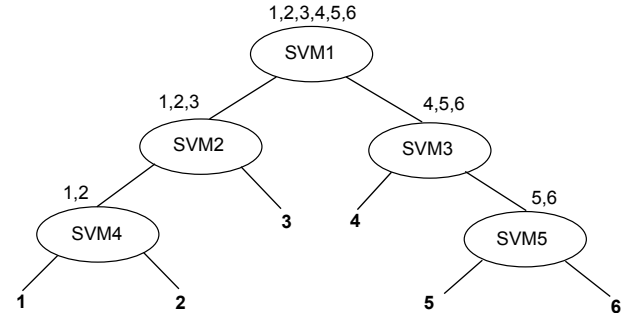


Fig. 2. SVM binary tree architecture

computation then one-against-one and one-against-rest methods. Classification procedure of the SVM-BT is based on a *coarse-to-fine* strategy that allows us to make a discrimination on coarse classes at the top of the decision tree architecture. A discriminative problem is solved gradually by stepwise classification. Thus, the coarse classification that separates two easy to differentiate classes, i.e. speech and non-speech, is performed at the beginning of the classification process. Classification problem results usually in high accuracy. Then the stepwise classification is made, until the one of leaves that represent the fine-gained class is obtained.

IV. EXPERIMENTS AND RESULTS

All the experiments were conducted on KEMT-BN1 database which contains recordings from the Slovak TV broadcast audio streams [14]. The database consists of 188 audio recordings in PCM 16kHz 16bit mono format in total duration 65 hours and manually annotated metadata files. The evaluation (training) was performed by only 8 recordings, in total duration about 117 minutes of speech and 21 minutes of non-speech audio segments. Testing phase was performed by using 4 recordings, each of which contained TV news, different from those used in evaluation phase. The overall duration of test data was 79 minutes. All the evaluations within training phase were based on the assumption that the SVM needs only small set of data in order to preserve generalization ability and to avoid the problem of overfitting.

At first, we divided each audio recording into non-overlapping segments with length 1s by a rectangular window and then time domain features, such as RMS of STE and ZCR, were extracted within each segment. Segments were further divided into 40 audio frames, each with 50ms duration and 25ms overlap. This phase of pre-segmenting helped us to remove silent segments by defining threshold of the average short-time energy and ZCR. Frame-based features were then calculated for each frame and used for classification with SVM. Thus, both training and testing were evaluated by these frame-based features with total dimension 36. Tab. II presents a summary of used features with corresponding dimensions.

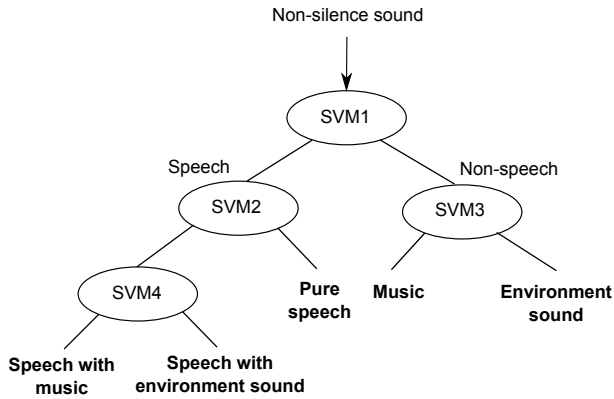


Fig. 3. SVM binary tree architecture adapted to BN audio data

Classification of the audio data, represented by feature vectors, was realized by the SVM-BT architecture, listed in Fig. 3. We used radial basis kernel and 5-fold cross-validation as evaluation functions for each SVM model, in order to get the best (optimal) parameters for the SVM and final fscore and accuracy of selected training data (Tab. III). Fscore evaluation parameter was computed as follows:

$$F - score = \frac{2 * Precision * Recall}{Precision + Recall}, \quad (11)$$

where *Precision* is defined as the proportion of the true positives against all the positive results and *Recall* shows the proportion of the true positives against the positive and false negative results. The resulting accuracy of the SVM-BT architecture was measured as the number of correctly predicted data to total test data, as shows Tab. III.

All experiments were performed by LIBSVM software¹ and available tools, which cooperate with this software.

TABLE II
USED FEATURES

Feature (descriptor)	MFCCs	ZCR	ASF	ROF	ASC	ASS
Dimension	13	1	19	1	1	1

TABLE III
CLASSIFICATION RESULTS FOR EACH SVM MODEL

SVM model	Best parameters	F-score [%]	Acc [%]
SVM1	C=32, gamma=0.125	97.42	95.08
SVM2	C=8, gamma=0.5	81.23	73.73
SVM3	C=0.25, gamma=0.125	87.96	78.52
SVM4	C=0.25, gamma=0.125	99.78	99.56

V. CONCLUSION AND FUTURE WORK

Multi-class audio classification system using SVM binary tree architecture was introduced in this paper. This system was adapted in order to discriminate between five types of acoustic events, with the highest occurrence in BN audio stream: pure speech, speech with music, speech with environment sound, environment sound and music. We used coarse-to-fine strategy as the classification method. Results in the Tab. III refer to the classification ability of each SVM binary classifier, which achieved relatively high values in all cases. It follows that used classification architecture is well

designed and convenient to solve multi-class discrimination problem. Used architecture address only discrimination ability of individual SVM classifiers in training phase. During testing phase may arise miss-classification error, caused by miss-classification of each previous SVM classifier in coarse-to-fine method. Therefore the classification accuracy of used system strongly depends on discrimination ability of the first SVM classifier: speech/nonspeech. We removed silent segments in pre-segmentation phase and also excluded speech segments with a low quality in order to decrease training time and computational complexity and increase separation ability of each SVM classifier.

Future work will be led to propose testing algorithm with minimum miss-classification error and to design an effective multi-class segmentation algorithm in combination with SVM-BT architecture that address this paper.

ACKNOWLEDGMENT

The research presented in this paper was supported by the Ministry of Education under the research project VEGA 1/0386/12 (30%) and Research and Development Operational Program funded by the ERDF under the projects ITMS-26220120030 (35%) and ITMS-26220220155 (35%).

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, *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2005, vol. 3497 (II), ch. Automatic News Audio Classification Based on Selective Ensemble SVMs, pp. 363–368. [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 radio broadcast content in radio devices," *International Conference on Consumer Electronics*, vol. art. no. 5012179, 2009.
- [5] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification," in *Proc. SIGIR*, 2003, pp. 282–289.
- [6] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multi-feature speech/music discriminator," in *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)*, ser. ICASSP '97, vol. 2. Washington, DC, USA: IEEE Computer Society, 1997, pp. 1331–1334.
- [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.
- [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] E. Vozarikova, J. Juhar, and A. Cizmar, "Acoustic events detection using mfcc and mpeg-7 descriptors," in *Communications in computer and information science: Multimedia Communications, Services and Security*, 2011, vol. 149, pp. 191–197.
- [11] H.-G. Kim, N. Moreau, and T. Sikora, *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. John Wiley & Sons, 2005.
- [12] C. K. On and P. M. Pandiyan, "Mel-frequency cepstral coefficient analysis in speech recognition," *Computing & Informatics 2006, ICOI'06*, no. 2, pp. 2–6, 2006.
- [13] S. Abe, *Support Vector Machines for Pattern Classification (Advances in Pattern Recognition)*. Springer-Verlang New York, Inc., 2005.
- [14] M. Pleva, J. Juhár, and A. Čizmar, "Slovak broadcast news speech corpus for automatic speech recognition," in *Proc. of the 8th Intl. Conf. on Research in Telecommunication Technology, RTT'07*, Liptovský Ján, Slovak Republic, Sep. 2007, p. 4.

¹<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

Automated Refactoring algorithm - AURA

¹Ivan HALUPKA (1st year), ²Emília PIETRIKOVÁ (2nd year)

Supervisor: ³Ján KOLLÁR

^{1,2,3}Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹ivan.halupka@tuke.sk, ²emilia.pietrikova@tuke.sk

Abstract—This article presents our research in the field of automated task-driven grammar refactoring. It is mainly concerned with our proposal of evolutionary algorithm, designed for automated refactoring of context-free grammars on the basis of transformation task. It also briefly introduces its reader to processes occurring in grammar refactoring, method for describing desired properties that refactored grammar should fulfill and argues about overall significance of grammar refactoring. Proposed algorithm transforms grammars in relatively large count of small and iterative steps by applying sequence of simple transformation procedures on initial grammar. This approach facilitates algorithms universality, and correctness of resulting grammars.

Keywords—Grammar refactoring, evolutionary algorithm, refactoring processes, task-driven transformation

I. INTRODUCTION

Our work in the field of automated grammar refactoring derives from the fact that two or more equivalent context-free grammars may have different form. Although two equivalent grammars generate same language, they do not necessarily share some other specific properties, measurable by grammar metrics [1]. Form in which context-free grammar is written may have strong impact on many aspects of its future application. For example it may affect general performances of parser used to recognize language generated by mentioned grammar [2], or it may influence and in many cases limit our choice of parser generator used to implement syntactic analyzer [2].

Since there is a close relation between forms in which grammar is expressed and the purpose for which grammar is being designed, different grammars generating same language become domain-specific formalizations of language. Because of this of this fact, ability to transform one grammar to another equivalent grammar actually becomes capability to shift between domains of grammars possible application. Although this ability makes each context-free grammar more universal in the scope of its application, its practical advantages may be easily overwhelmed by difficulties that this approach can cause. Problem is that manual grammar refactoring is in many cases a non-trivial and error-prone task, especially in case of larger grammars. This is an issue because in general there is no formal way of proving that two context-free grammars generate same language, since this problem is undecidable.

In our work, we address this issue by proposing evolutionary algorithm for automated task-driven grammar refactoring. Algorithm is called *AURA*. Main idea behind our algorithm is to apply sequence of simple transformation processes on the chosen context-free grammar in order to produce equivalent

grammar with desired properties. In current state of development algorithm requires that grammar's production rules be expressed in BNF notation. These refactoring processes are more closely discussed in section III, while refactoring algorithm itself is discussed in section V. Desired properties of grammar produced by algorithm are defined by objective function, which we discuss in section IV.

II. MOTIVATION

Grammarware engineering is an emerging discipline of software engineering, which aims to solve many issues considering grammar development and promises overall rise in quality of produced grammars and productivity of their development [3]. Grammar refactoring is a process that may occur in many fields of grammarware engineering, such as grammar recovery, evolution or customization [3]. In fact, it is one of five core processes occurring in grammar evolution, beside grammar extension, restriction, error correction and recovery [4]. Problem is that unlike program refactoring which is well established practice, grammar refactoring is weakly understood and hardly practiced [3].

If there is a one clear purpose for which grammar is being developed, its specification for experienced grammar engineer is usually not an issue. Problem arises when grammar is being developed for multiple purposes [4], or grammar engineer lacks of knowledge about grammar's future purpose. In the first case problem is usually solved by developing multiple grammars of one language [4]. This need for development of multiple grammars could be replaced by developing only one grammar generating given language and automatically refactoring it to another form suited to satisfy certain requirements, thus increasing grammar engineer's productivity. Actually, this is one of the main objectives of our work in the field of grammar refactoring.

In case when grammar engineer lacks of knowledge about some aspect of grammar's future purpose, its final shape may not satisfy some of the specific requirements, even if it generates correct language. In this case grammar must be either refactored or rewritten from scratch thus draining another of valuable resources. Automated or even semi-automated way of refactoring mentioned grammar could produce significant savings in this redundant resource consumption. This two scenarios are not the only where efficient refactoring tool is needed, in fact it can be useful in all cases where we have a grammar whose form needs to be changed with preserving language that it generates. In this case we see two main domains of our algorithms application which are *adaptation of legacy grammars* and *grammar inference*.

Fact is that parser generators and other implementation platforms for context-free grammars develop over time. Newly established platforms and other tools operating with context-free grammars may require different form in which grammar should be expressed than tools of previous technological generation, or operate with unequal efficiency over the same grammar forms. Kent Beck states that programs have two kinds of value: what they can do for today and what they can do for tomorrow [5]. When we take this principle into the account, we can say that ability to refactor context-free grammar in order to adjust it to requirements of current platforms is in fact capability to add value to language's legacy formalization.

Grammar inference is defined as recovering grammar from a set of positive and negative language samples [6]. Considering inferred grammar, it mostly focuses on resolving issues with over-generalization and over-specialization of generated language [7], while grammar's form stays behind as a secondary concern. Grammar recovery tools in general do not allow its users enough fine-grained tuning options of recovered grammar's form, making it in many cases difficult to comprehend, and not useful until refactored [8].

III. REFACTORING PROCESSES

In our approach we use grammar refactoring processes as a only tool of incremental grammar refactoring. Formally, grammar refactoring process is a function that takes some context-free grammar $G = (N, T, R, S)$ and on its basis creates a new grammar $G' = (N', T', R', S')$ equivalent to grammar G . Mentioned function may also require some additional arguments, known as process parameters. To each assignment of actual values to required process parameters of specific grammar refactoring process we refer as refactoring process instantiation, and specific grammar refactoring process with assigned actual values of its required process parameters we call instance of this refactoring process.

At this stage of development, we have experimented with the base of eight grammar refactoring processes (Unfold, Fold, Remove, Pack, Extend, Reduce, Split and Nop), from which first three are adopted from Ralf Lämmel's article about grammar adaptation[9], and others are proposed by us. For the purpose of better understanding what refactoring processes really are and how they work, in following subsections we briefly introduce two of mentioned, namely Pack and Nop.

In our research we tend to keep the process base as small as possible, and contained refactoring processes as universal as possible. This is mainly because with growth of refactoring process base, state space of possible solution grammars also grows and thus size of process base has a significant impact on calculation complexity of algorithm. Lack of domain-specific refactoring processes is on the other half compensated by overall openness of process base, which makes it's expansion or reduction a relatively trivial task. In fact only refactoring process required by algorithm, that must at all time reside in process base is process Nop.

A. Nop

Nop or identical transformation is grammar refactoring process that maps context-free grammar G into same context-free grammar G , or in other words, it does not impose any changes on the grammar.

B. Pack

Pack is grammar refactoring process that creates transformed grammar G' on the basis of three process parameters, which are mandatory parameter called packed production rule (Pr), and optional parameters initial package symbol (Ps) and package length (Pl). These parameters must have following properties:

$$Pr \in R \quad (1)$$

$$Ps \in \mathbb{N} \wedge Ps \geq 0 \wedge Ps < rightSideLength(Pr) \quad (2)$$

$$Pl \in \mathbb{N} \wedge Pl > 0 \wedge Pl \leq rightSideLength(Pr) - Ps \quad (3)$$

In case when initial package symbol is not defined we assume that $Ps = 0$. If package length is not defined we define it as $Pl = rightSideLength(Pr) - Ps$. Function *rightSideLength* returns number of symbols contained within right side of production rule.

Pack replaces some sequence of symbols contained within right side of packed production rule with new non-terminal and creates rule whose left side is this new non-terminal and right side is mentioned sequence of symbols. This sequence of symbols is defined by initial package symbol and package length. More precisely it is a sequence of Pl symbols starting from symbol whose position within packed rule is $Ps + 1$.

IV. OBJECTIVE FUNCTION

We adopt somewhat modified understanding and notation of objective functions from mathematical optimization. In this case objective function describes properties of context-free grammar that one seeks to achieve by refactoring. On the other hand, it does not describe way in which refactoring should be done, and condition in which desired grammar's properties are achieved.

In our view the objective function consists of two parts which are *objective* and *state function*. Our automated refactoring algorithm works with only two kinds of objectives, which are minimization and maximization of state function. We define state function as an arithmetic expression, whose only variables are grammar metrics, calculable for arbitrary context free grammar. As such state function represents a tool for qualitative comparison of two or more equivalent context-free grammars.

Until now we have experimented with some grammar size metrics, like number of non-terminals (*var*) and Halstead effort (*hal*) [1]. Example of objective function defining refactoring task to be performed on the grammar G executable by our algorithm is:

$$f(G) = minimize \ 2 * var + hal \quad (4)$$

V. REFACTORING ALGORITHM

Main idea behind our grammar refactoring algorithm is to apply sequence of grammar refactoring processes on chosen context free grammar, in order to produce equivalent grammar with lesser value of objective function in the case of minimization objective, or higher value of objective function in the case of maximization. Since it is evolutionary algorithm, in order to be executed it also requires some other input parameters, except *initial grammar* and *objective function*. Algorithm requires three other input parameters, which are *number of evolution cycles*, *population size* and *life length of generation*. First two of these parameters are characteristic for

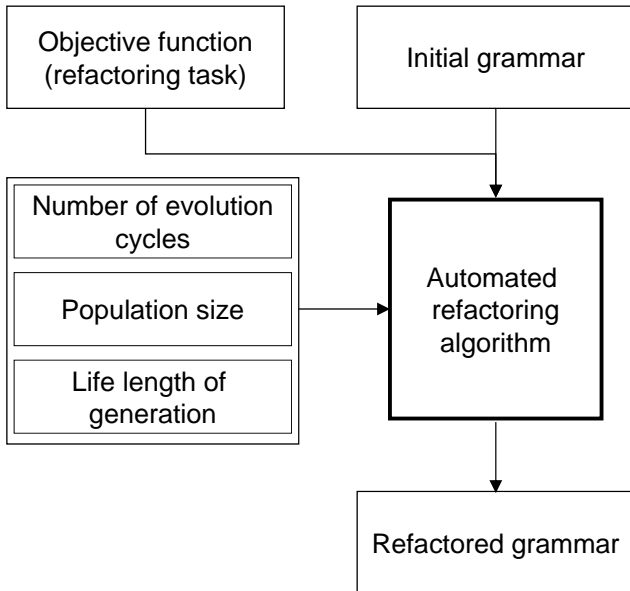


Fig. 1. Black-box view of Automated refactoring algorithm

algorithms of similar type, while third is our own. Black-box view of our evolutionary algorithm is also shown on Fig. 1.

As can be seen on Fig. 2 presenting the white-box view of our algorithm, central figure of AURA is an abstraction called *population of grammars*. In our view population of grammars is set containing constant number of grammar population entities. Its main property is that after performing arbitrary step of our algorithm number of elements in population of grammars is always equal to population size.

Further we define grammar population entity as arranged triple of elements which are *post-grammar*, *process chain of grammar generation* and *objective functions difference*. Post-grammar is context-free grammar equivalent to initial grammar. Process chain of grammar generation is a sequence of refactoring process instances which was used to create post-grammar, from the corresponding post-grammar of previous generation. Number of refactoring process instances in each process chain of grammar generation is always equal to life length of generation. Objective functions difference is difference between values of objective function calculated for post-grammar of current and corresponding post-grammar of previous population of grammars.

A. Refactoring process instantiation

All process instances occurring in our algorithm are created automatically in one of three procedures, which are called *random process creation*, *random parameters creation* and *identical process creation*.

Random process creation creates instance of random refactoring process with random parameters. First step of this procedure randomly selects process from the base of grammar refactoring processes. In this procedure each grammar refactoring process has a same probability that it will be selected. The second step of procedure defines concrete process parameters for this process on the basis of grammar on which process instance will be applied. All possible combinations of process parameters that respect restrictions defined by specific refactoring process have same probability that they will be generated in this procedure.

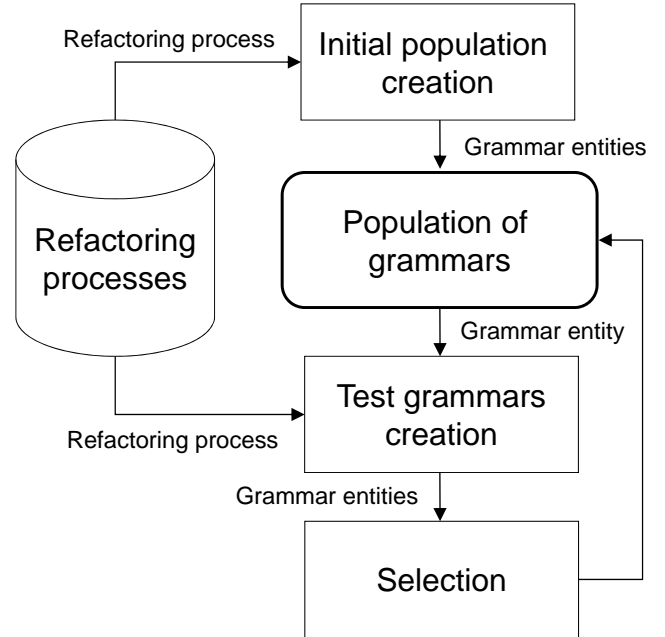


Fig. 2. White-box view of Automated refactoring algorithm

Random parameters creation creates process instance originating from some other process instance. Two mentioned process instances share the same refactoring process, but their process parameters may differ, since new process parameters have been created in procedure analogous to second step of random process creation procedure. Only exception from this rule occurs in case when there is no acceptable combination of process parameters for given refactoring process in order to be applicable on given context-free grammar, in which case random parameters creation procedure returns instance of Nop refactoring process.

Identical process creation creates instance of Nop grammar refactoring process.

B. Initial population creation

In the first phase of Automated Refactoring Algorithm initial population of grammars is being created, and as such this phase is not being repeated throughout the algorithm.

First step of this phase is creating process chains of grammar generation, for each grammar population entity. All process instances of each process chain created in this phase of algorithm are created in random process creation procedure, except one whose all processes are created in identical process creation procedure. Reason for this exception is to guarantee that initial grammar will be incorporated in the initial population of grammars. Since sequence of process instances contained in process chain must be applicable on grammar for which are they being generated in exact order, we must consider all changes on grammar done by one refactoring process instance in order to be able to generate next process instance of process chain. We solve this issue by generating intermediate grammars after each random process creation procedure by applying this refactoring process instance on grammar for which random refactoring process instance is being generated, and then generate next random process instance of process chain on the basis of mentioned intermediate grammar. In order to better understand idea behind this approach, we provide example of creating process chain consisting of three random

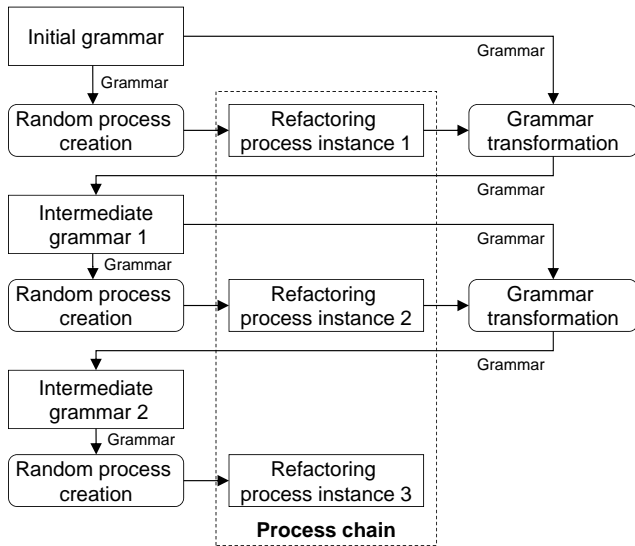


Fig. 3. Creation of random process chain

refactoring processes for initial grammar. This example is showed on Fig. 3.

Second step of the first phase of algorithm creates corresponding post-grammars for each grammar population entity by applying its process chain on initial grammar, and finally third step calculates difference of objective function calculated for initial grammar and post-grammar of corresponding grammar population entity.

C. Test grammars creation

Second and third phase of algorithm, called test-grammars creation and selection are being repeated in sequence for number of evolution cycles. In test-grammars creation, we create three test grammar population entities for each grammar population entity. These entities are called *Self-test grammar*, *foreign-test grammar* and *random-test grammar*.

Self-test grammar is created on the basis of corresponding grammar population entity and process chain, generated on the basis of this entity's process chain. All refactoring process instances in the newly generated process chain are created in random parameters creation procedure, and algorithm of their creation is analogous to algorithm of creating random process chain in the initial population of grammars creation phase. So self-test grammar is grammar population entity, containing grammar which was created on the basis of same refactoring processes on which original, tested grammar was created, but this processes may have different process parameters.

Foreign-test grammar is created in the similar procedure as self-test grammar, with exception that new population entity is not being created on the basis of tested grammars process chain, but on the basis of some other grammar population entities process chain. This population entity is randomly selected from the population of grammars.

Random-test grammar is created in the procedure analogous to procedure of creating random grammar population entity in the first phase of algorithm, with the exception, that random process chain is not being generated for initial grammar, but for grammar contained within tested grammar population entity.

D. Selection and evaluation

In the selection phase of algorithm we compare objective function's value of each grammar within the population of grammars, with objective function's values of corresponding test grammars, and choose grammar with best value of objective function. This is grammar which will be incorporated in the next generation of population of grammars. In case when chosen grammar is tested grammar, no changes occur and corresponding grammar population entity is preserved in population of grammars, while in opposite case tested grammar population entity is removed from population of grammars and substituted by test grammar population entity with best value of objective function.

Fourth and final phase of algorithm is performed after all evolution cycles have ended. In this phase we compare values of objective function calculated for each grammar within the population of grammars and choose one with highest or lowest value, depending on the objective. This is solution grammar and as such result of automated refactoring.

VI. CONCLUSION

In this article we presented our algorithm for automated refactoring of context-free grammars. Its main advantage is relatively broad scope of its application, while its main disadvantage is relatively high computational complexity. In future we would like to focus on task of lowering its computational complexity, and we would like to expand it in a way that enables the user to define what we call grammar invariants. Grammar invariants are native grammars properties which are required to be preserved in the refactored grammar. This would enable algorithm's user to influence form of resulting grammar on a greater scale, and guarantee that refactored grammar will satisfy expectations.

ACKNOWLEDGMENT

The work presented in this paper was supported by SRDA Project of Slovak-Slovenian Research and Development No. SK-SI-0003-10 "Language Patterns in Domain-specific Languages Evolution".

REFERENCES

- [1] J. Cervelle, M. Crepinsek, R. Forax, T. Kosar, M. Mernik, and G. Roussel, "On defining quality based grammar metrics," in *Computer Science and Information Technology, 2009. IMCSIT '09. International Multiconference on*, 2009, pp. 651–658.
- [2] T. Mogensen, *Basics of Compiler Design*. Copenhagen, Denmark: University of Copenhagen, 2010. [Online]. Available: http://www.diku.dk/hjemmesider/ansatte/torbenm/Basics/basics_lulu2.pdf
- [3] P. Klint, R. Lämmel, and C. Verhoef, "Toward an engineering discipline for grammarware," *ACM Transactions on Software Engineering Methodology*, vol. 14, no. 3, pp. 331–380, 2005.
- [4] T. L. Alves and J. Visser, "A case study in grammar engineering," in *Proceedings of SLE'2008*, 2008, pp. 285–304.
- [5] M. Fowler, *Refactoring: Improving the Design of Existing Code*. Boston, MA, USA: Addison-Wesley, 1999.
- [6] M. Mernik, D. Hrnčić, B. Bryant, A. Sprague, J. Gray, Q. Liu, and F. Javed, "Grammar inference algorithms and applications in software engineering," in *Information, Communication and Automation Technologies, 2009. ICAT 2009. XXII International Symposium on*, 2009, pp. 1–7.
- [7] A. D'Ulizia, F. Ferri, and P. Grifoni, "A learning algorithm for multimodal grammar inference," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 41, no. 6, pp. 1495–1510, 2011.
- [8] N. Kraft, E. Duffy, and B. Malloy, "Grammar recovery from parse trees and metrics-guided grammar refactoring," *Software Engineering, IEEE Transactions on*, vol. 35, no. 6, pp. 780–794, 2009.
- [9] R. Lämmel, "Grammar adaptation," in *Proc. Formal Methods Europe (FME) 2001*, 2001, pp. 550–570.

Automatic concept identification in slovak texts

¹Viliam Ročkai (5th year), ²Marián Mach

Supervisor: ²Marián Mach

^{1,2}Department of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹viliam.rockai@gmail.com, ²marian.mach@tuke.sk

Abstract—WordNet dictionaries are a commonly used resource in the NLP field. Many papers discuss the theme of WordNet dictionaries, but the processes devoted to their automated construction have still much space for improvement. Many of them are based on machine translation strategies converting the dictionaries from one language to another. Most of them have one attribute in common: they use the knowledge about the language used in the form of dictionaries or even simple grammar parsers. In WordNet dictionaries, terms are represented in the form of concept hierarchies (hypernyms, hyponyms, ...). Since the ALOC (associative learning of concepts) approach was used to achieve similar concept structures, we can assume that it could also be used in the area of automated slovak WordNet dictionary construction. This paper discusses such application of ALOC, in the form of correct assignment of new concepts into the existing WordNet hierarchy.

Keywords—ALOC, concept, NLP, similarity, slovak language

I. INTRODUCTION

This paper represents the state of art in obtaining a set of words related to a set of input words, thus constructing a semantically closed term set, using associative learning of concepts (ALOC) [4]. In [9], a method for computing word relatedness using ALOC was presented. In this paper, we present a method extending and unifying previous work. Identifying similar entities is usually supported by some sort of background knowledge in the form of ontologies [7], existing WordNet dictionaries [1] [6] or named entity recognition (NER) systems [5]. Our model does not contain any background knowledge. Using just three words as the input, the method is able to construct a wide set of related concepts based on the nature of the relation existing between the words in this word triplet.

II. LEARNING PROCESS

Confabulation theory offers a comprehensive detailed explanation of the mechanism of thought. As described in [3], it could be used as a NLP approach for automatic phrase recognition and completion. Our implementation [4] is based on such approach and works with symbols that are given by a set of excited neurons in some cortical region due to confabulation theory. An invariant sensory input and can be seen as an attribute of the perceived object. In this theory, symbols are defined as the representation of these attributes. A concept is defined as a symbol with its associated neighbourhood. Thus, the same concept can be activated by different sets of symbols. Associations anchor these symbols to each other. The establishment of an association between two symbols depends

on their co-occurrence significance. The significance can be determined by utilizing the following formula:

$$sig_i(a, b) = \frac{p_i(a, b)}{p(a)p(b)}, \quad (1)$$

where a and b are discrete random variables and i is their contextual distance. The value $sig_i(a, b)$ stands for mutual significance and is based on the information theory formula of mutual information. In our case, using a natural language corpus as the source, we can understand $p(a)$ and $p(b)$ as the probabilities of encountering words a or b in the text. The distance between these two words is their contextual distance and is proportional to the number of words located between them (starting with the number 1 for immediately adjacent words, so that the value -1 can stand for inverse direction). The value $p_i(a, b)$ then stands for the probability of seeing these two words together (in the particular contextual distance i). The mutual significance is defined as the ratio of mutual probability of symbols (discrete variables) a , b to their prior probabilities [8]. If a and b are independent, then $sig_i(a, b) = 1$. We assume an existing association in the case when $sig_i(a, b) > threshold$ – the variables a and b are not independent if $threshold$ is greater than 1.

A. Lexicon

During the first phase of learning process, a lexicon of known symbols is built. In this paper, a symbol is understood to be a single word or a punctuation mark from the input text, a so-called token. According to [4], a lexicon consisting of 3500 distinct tokens should cover more than 85% of the whole size of corpus written in English language. The size of corpus used was approximately 200 MB (consisting of cca 45 000 000 words). The corpus used in this paper consists of a huge number of articles available in the plain text format, written in Slovak. The articles originate from the Slovak Wikipedia project. In the experiments below, a lexicon of 6000 (most frequent, including stop-words) symbols was used, nearly twice the size compared to the research before – this was chosen to cover more of the commonly used Slovak words, which are needed for the examples presented.

B. Knowledge fascicles

Fascicles are implemented as matrices storing the values of co-occurrences of symbols contained in the lexicon (Figure 1). These numbers are later used for computing significance and weight values. The number of fascicles equals the number of different contextual distances – each fascicle represents an

	window	he	is	was	a	sword	brave	boy
window			3	17				
he		1	8	13				
is					23		1	
was					2			
a	5					2	1	9
sword				4				
brave								3
boy			5	7				

F
A

Fig. 1. Fascicles are implemented as matrices storing the values of co-occurrences of symbols contained in a lexicon.

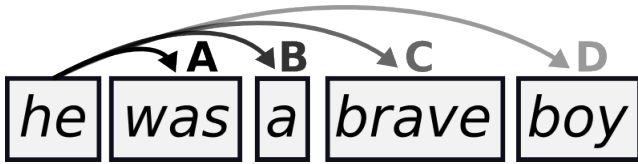


Fig. 2. Learning process is performed over a stream of tokens. In this case, co-occurrences are stored in 4 separate matrices (A, B, C, D), called fascicles.

individual contextual distance. In the experiments below, learning was carried out using 5-word windows, thus 4 contextual distances were covered and 4 fascicles were used.

The learning process consist of storing co-occurrences of token (word) pairs for four contextual distances. Learning is done over a stream of tokens (i.e. English literature). It consists of two main phases.

In the first phase, a lexicon of a fixed number of tokens is generated. Then, in the second phase, associations between those tokens are induced in a number of contextual distances (Figure 2). The storage process of these word co-occurrences uses a token window. This token window shifts forward through the token stream. The window is trimmed from the right, so that all tokens in the window are known (can be found in the lexicon). That means, if the fifth token in the window is unknown symbol, and all the others are know, only the left part consisting of 4 tokens is used as a token window. The co-occurrences of all contextual distances in this trimmed window are stored. These co-occurrences are then used for computing the probabilities used below. For deeper understanding, more information about the learning process can be found in [4].

III. SIMILARITY COMPUTING

A. Simple similarity computing

The relatedness strength of a token pair in some contextual distance is defined according to this formula based on the Jaccard index [2]:

$$Rel_i(a, b) = \frac{|O_i(a) \cap O_i(b)|}{|O_i(a) \cup O_i(b)|}, \quad (2)$$

where $O_i(a)$, $O_i(b)$ stand for the semantic neighbourhood of symbols a and b in a contextual distance i . Semantic

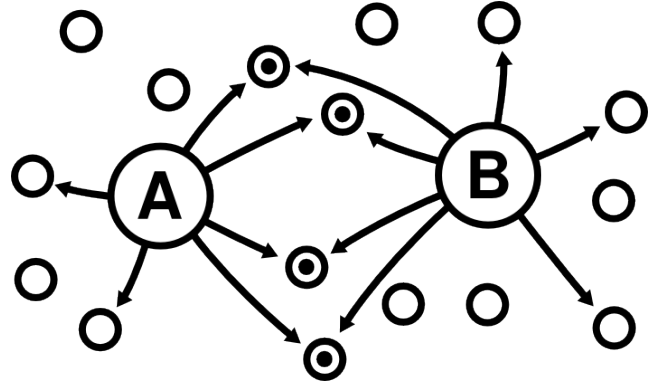


Fig. 3. Tokens are associated with each other. The number of tokens associated with both tokens A and B is directly proportional to the relatedness strength of the pair A-B.

neighbourhood $O_d(x)$ is defined as the set of all tokens associated with the token x in the contextual distance d . If $|O_i(a) \cap O_i(b)| = 0$ the tokens are not related and their relatedness strength is 0. If the value of $Rel_i(a, b)$ is greater than 0, tokens are somehow related to each other. Mutual associations of a token pair in one contextual distance are depicted in Figure 1. The semantic neighbourhood of the token A is associated with 6 tokens. The semantic neighbourhood of the token B is associated with 7 tokens. They both share 4 associated tokens and the number of distinct tokens associated with both symbols is 9. Thus $Rel_i(A, B) = 4/9 = 0.44$.

When relatedness strength is being calculated through several contextual distances, we can assign a weight for each distance. The formula for computing weighted relatedness strength for several contextual distances is defined below:

$$Rel_w(a, b) = \sum_{i=1}^n w_i Rel_i(a, b), \quad (3)$$

where $Rel_i(a, b)$ stands for the relatedness strength of a token pair in contextual distance i , n stands for the number of contextual distances and w_i stands for the weight of the relatedness strength for the contextual distance i .

IV. CONCEPT CLUSTERING

Computing the relatedness of some input token with all other tokens from the lexicon leads to the construction of a long list of tokens ordered by their relatedness value. The main issue with such a list is the determination of some kind of a strict threshold, enabling to separate relevant related tokens from the others. The list illustrates the functionality of ALOC to identify relevant words to the input word, but there are some issues:

- List length – the list is usually shorter than the lexicon by a few words only.
- Errors in symbol ordering – it is common that symbols are not ordered in line with our expectations, based on empirical knowledge of the language.
- Word-sense disambiguation – the word *May* could mean either the *month* or the verb *could*.

The symbol clustering technique introduced in this section is designed to tackle these issues. First of all, let us define a symbol cluster. When computing the ordered list for some symbol, a single symbol is used as the input. This symbol could be

part of many sense sets, synsets which are connected through some type of WordNet relation. Typical example of a valid sense set could be the sister-term relation in WordNet. While the word *January* is found only in single WordNet synset, the word *March* is found in 14 distinct synsets (with meanings including a *month*, a *genre of music*, *steady advance*, etc.). In WordNet, synset is defined by enumerating its elements and by a short gloss, describing the synset in a natural language. Since the ALOC approach assumes no understanding of the particular natural language, the only way to define a cluster of symbols is by enumerating some of its elements. While the word *March* is a part of many synsets, the word couple *March* and *January* are elements of synsets which are related through the sister-term relation, sharing a common parent through the direct hyponym relation. These two words would be sufficient to clearly identify a cluster denoting months. By combining two or more ordered lists of related symbols, it could be possible to automate the generation of some word sense sets used in WordNet. We assume that all words contained in a single synset of size bigger than two would be related to each other with approximately the same value of relatedness. For better understanding and the ability to visualise the task, the symbol semantic distance is introduced:

$$D_w(a, b) = 1 - R_w(a, b), \quad (4)$$

where $D_w(a, b)$ is the value of semantic distance between symbols a and b and $R_w(a, b)$ denotes relatedness strength of such symbol pair. Each list of tokens, ordered by the relatedness strength to some input token could be recomputed utilising the formula above to use the distance measure instead of relatedness measure.

The knowledge of a single list of related tokens ordered by semantic distance is not enough to define a cluster of symbols which would denote a single clearly defined semantic set. Since the 2-combinations of all symbols within such a set are expected to have approximately same distances, this could be used as an imaginary border of this set. The semantic area is, in this paper, defined as the enumeration of its elements:

$$M_c = \{s_1, s_2, \dots, s_n\}, \quad (5)$$

where M_c is a set denoting the semantic area, symbols s_1, s_2, \dots, s_n are its elements and n is the number of elements. The condition defining whether some new, unclassified symbol t belongs into M_c is:

$$\frac{1}{n} \sum_i^n D_w(s_i, t) \leq \frac{1}{C^2(n)} \sum_i^n \sum_{j=i+1}^n D_w(s_i, s_j), \quad (6)$$

where i and j are the elements of the semantic area M_c and $C^2(n)$ is the number of 2-combinations from M_c :

$$C^2(n) = \frac{n!}{2!(n-2)!} \quad (7)$$

The formula (6) could be easily interpreted as follows: a symbol t belongs to M_c only if the average semantic distance of the symbol t to all symbols from M_c is not greater than the average semantic distance between all symbol pairs from M_c .

TABLE I
THE OUTPUT OF ALOC CLUSTERING METHOD FOR AN INPUT WORD TRIPLET CONSISTING OF THREE MONTHS.

input	január, február, marec
output	január, február, marec, apríl, december, júl

TABLE II
THE OUTPUT OF ALOC CLUSTERING METHOD FOR AN INPUT WORD TRIPLET CONSISTING OF THREE FIRST NAMES.

input	peter, michal, marek
output	peter, michal, marek, daniel, juraj, václav, richard, viliam, anton, jozef, andrej, Ľudovít, eduard, miroslav, ivan, pavol, jakub, rudolf, ladislav, Ondrej, ján, martin, tomáš, alexander, matej, františek, štefan

TABLE III
THE OUTPUT OF ALOC CLUSTERING METHOD FOR AN INPUT WORD TRIPLET CONSISTING OF THREE NUMBERS.

input	jeden, dva, tri
output	jeden, dva, tri, štyri, dve

V. EXPERIMENTS

To demonstrate the functionality of our concept clustering approach, the semantic area which will serve as the input has to be chosen in a specific way. Its elements should unambiguously denote a single semantic class and they must be frequent in the language, so that they are part of the lexicon. The size of the input semantic area was 3 symbols, since in such set the number of distances between all member symbols (the cluster consistency) is the same as the number of distances between the new symbol and each member symbol (symbol to cluster relevancy).

For these purposes, sets like *months of a year*, *first names*, or *numbers* could be used. The input consists of three member subsets from such semantic sets. The expected output is the discovery of new symbols from full semantic sets (i.e. enlarging used three member subsets by other symbols from chosen input semantic areas).

All texts from the Slovak Wikipedia project were used for the ALOC association induction phase. The total size of the plain-text documents used was 121 MB. The lexicon consisted of 6000 most frequent words from those documents. Associations were accumulated in four contextual distances (1, 2, 3, 4), with a threshold value set to 2. All weights used in the relatedness strength formula had the value $1/4$. The outputs for several inputs under these conditions are shown in tables I, II and III. All results have 100 % precision.

In the next experiment, we created all possible (220) triplet combinations from the set of twelve months. Each triplet was used as an inputs to the clustering algorithm and for each of them a cluster of words was found. In all result clusters, the new words discovered were again the words naming months. The full set of twelve months was found in 10 % cases. In the remaining cases, the size of resulting cluster was between 3 (which contained the input triplet only – no new symbol was discovered) and 11. The histogram depicted in Fig. 4 illustrates the recall achieved in this experiment. The axis x represents all possible recall values while the y axis shows the number of triplets for which the given recall value has been obtained. The average recall for all 220 triplets was 0.66 and the precision achieved for each triplet was 100 %.

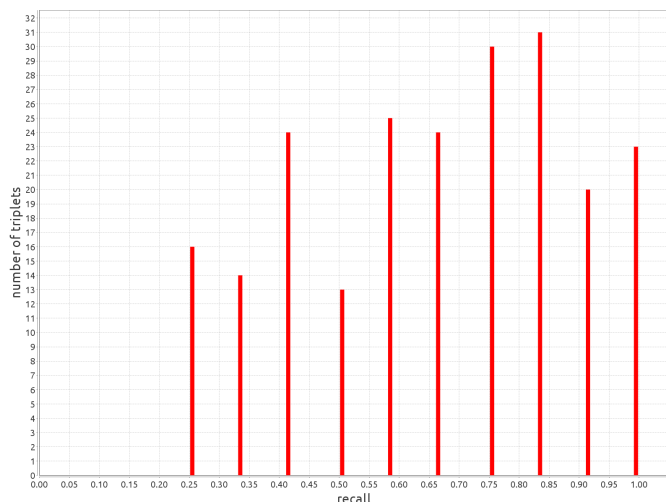


Fig. 4. Recall value histogram for all combinations of month triplets.

VI. CONCLUSION

In this paper, we explored the utility of the ALOC approach in clustering semantically similar words based on their semantic distance and gave a concise description of the methods we used. To illustrate the successfulness of the clustering method, we used it to generate semantic sets from their three-member subsets. The input consisted of word triplets selected from WordNet sister synsets (denoting months, first names, numbers) and we tried to find other words extending input triplets to the full synsets. This was accomplished with 100% precision and very satisfactory recall. The approach was presented on a very restricted domain of examples, which were all in Slovak. The approach was tested with other (Czech, English, Korean) languages and was achieving analogous results. Further research should be focused on several areas. Extending the domain of inputs and languages is needed for more precise evaluation of the current state of the approach. Modifying the ALOC implementation to be able to handle unlimited lexicon size would make this evaluation even more objective. The last area is to modify the ALOC approach in a way, that it would be able to process word pairs and n-grams in a similar way how it is processing single words in current state.

CORPUS SAMPLES

... Toto dlhé súhvezdie vinúce sa od Orióna hlboko na juh už oddávna pripomínalo ľuďom riek. Egypťania v ňom videli obraz posvätného Nílu, antickí Európania pravdepodobne zase severotalienskú riek Pád. S Pádcom ho stotožňuje aj antická báj o Faethonovi, synovi boha Slnka Helia. Faethon raz uprosil otca, aby mu požičal jeho slnečný voz. Márne ho Helios odhovárал, že riadenie voza je ťažké a nebezpečné, Faethon trval na svojom ...

... V Paríži našiel Chopin umeleckú spoločnosť, rovnako ako príležitosť uplatniť svoj talent a stať sa celebritou a neskôr slušne zarábať výukou klavírnej hry. Spriatelil sa s osobnosťami slávnych mien: Hector Berlioz, Franz Liszt, Vincenzo Bellini, Ferdinand Hiller, Felix Mendelssohn, Heinrich Heine, Eugène Delacroix, kníže Adam Jerzy Czartoryski, Alfred de Vigny, Charles-Valentin Alkan ...

ACKNOWLEDGMENT

The work presented in this paper was supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic under grant No. 1/0685/12 (50%). This work is also the result of the project implementation Development of the Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (50%).

REFERENCES

- [1] Fellbaum C.: *WordNet: An Electronic Lexical Database*, The MIT Press, Cambridge, MA, 1998
- [2] Jaccard, Paul (1901), "Étude comparative de la distribution florale dans une portion des Alpes et des Jura", *Bulletin de la Société Vaudoise des Sciences Naturelles* 37: 547–579.
- [3] Nielsen R. H.: *Confabulation Theory: The Mechanism of Thought*, Springer, 2007
- [4] Ročkai V., 2005. *Mining of Concepts and Semantic Relations from Texts in Natural Language* Diplomová práca, Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky, Technická univerzita, Košice, 2005
- [5] Blume M.: Automatic Entity Disambiguation: Benefits to NER, Relation Extraction, Link Analysis, and Inference. In *Proceedings of the International Conference on Intelligence Analysis*, McLean, Virginia, USA, 2005.
- [6] Farreres X., Rigau G., Rodríguez H.: "Using WordNet for Building WordNets," *Proceedings of COLING-ACL Workshop on Usage of WordNet in Natural Language Processing Systems*, 1998.
- [7] Hassell J., Aleman-meza B., Arpinar I. B.: *Ontology-Driven Automatic Entity Disambiguation in Unstructured Text*, In *Proceedings of the 5th International Semantic Web Conference (ISWC)*, Volume 4273 of *Lecture Notes in Computer Science*, pages 44–57, Athens, GA, USA, 2006.
- [8] Ročkai V., Kende R.: *Associative learning of concepts*, AI METH 2007, CAMES, Gliwice, Poland 2007
- [9] Ročkai V., 2007. Automatické generovanie slov patriacich do jednej syntakticko-sémantickej triedy, WIKT 2007 - *Proceedings of 2nd Workshop on Intelligent and Knowledge oriented Technologies* (Košice, November 15 - 16, 2007). Centre for Information Technologies, FEI TU Košice, Slovakia, 2008

Basis for Model Oriented Language Composition

¹*Dominik LAKATOŠ (2nd year)*
Supervisor: ²Jaroslav PORUBÁN

^{1,2}Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹dominik.lakatos@tuke.sk, ²jaroslav.poruban@tuke.sk

Abstract—Composition of computer languages is one of the main processes in language evolution as it provides possibilities for creating new languages by merging. In this paper we present ideas of language composition with regard to language structure consisting of abstract syntax, concrete syntax and semantics. We find common problems and conflicts. Main composition patterns are used together with simple examples to explain possible trends in language composition and how they can be used to help with composition problem. Special attention in this problem is given to the domain-specific languages.

Keywords— composition conflicts, composition patterns, language composition, language design

I. INTRODUCTION

*“I should not design a small language,
 and I should not design a large one.
 I need to design a language that can grow.”*

Guy L. Steele Jr. [1]

The world around us has been evolving for centuries. Not just nature but people, too. Our understanding is changing with each discovery, our languages adapt to changes in our surrounding, and architecture styles shifts as people want to see something new. Nothing is constant. Computers are changing really fast in few years and are changing everything connected to them. The same should apply to computer languages, they need to adapt to always changing needs and expectations of users. Composition is one of the main processes for evolution, as an evolution is mostly consisting of changing existing things to the new and better ones. In this paper we will try to expose the idea of computer language composition as well as problems connected with it.

II. LANGUAGE COMPOSITION

Software system is usually composed of several sentences of one or more languages. For example Java application is composed of several classes written in separate files. It is very similar with almost any programming languages, complete application code is written in more than one file, which are then composed to get final working system. Actually modern software systems are written in more than one computer languages – connecting a general-purpose language with configuration languages, template languages, database query languages, graphical interface design languages, transformation languages and many more. It can be hard to

precisely specify which language is the main language of software system.

III. DIFFICULTIES OF COMPOSITION

Language composition is not a simple process of merging multiple languages. Problem with composition of languages is usually in the effort to combine different languages in a term of meaning and notation form.

Meaning of composition result should be somehow predictable. If user knows both combined languages in its native form, it would be great advantage if he would be able to understand meaning of a composed language created from both combined languages without any more explanations. So languages should not be composed with totally changed semantics. Concepts of language should be designed to connect logically, so its semantics are predictable.

Composition of two or more languages can be problematic due to existence of same concepts in both languages, in such case we have concept conflict. Conflict can be discovered in different language parts. There can be conflict only in the name of concepts but the meaning is completely different. For example concept *address* can be in one language used as a *postal address* and in the second language as a *network address*. In some composition cases it is probable to have to compose languages with identical concepts, which are defined with different connecting concepts or the scope of concept meaning can be different. Problem of conflicts is more extensively described in next section.

Another big problem for language designers is almost zero help in language design tools and language workbenches for actual composition of languages. So if there is need to reuse already defined language concepts and connection for next language it has to be done manually by copying and there is no reference to original language.

The main recognized problems of language composition are:

- different ways for composition of languages and its meaning,
- conflicts between composed languages,
- minimal existence of tool support for composition of languages.

IV. COMPOSITION PROBLEMS WITH REGARD TO LANGUAGE STRUCTURE

Internal structure of language can be separated into three main parts: abstract syntax, concrete syntax and semantics [2].

Composition of languages needs to be done on all three parts. Each part usually needs own composition technique and have own problems connected with composition.

A. Abstract syntax

Composition of different abstract syntaxes is not always a direct problem, mostly when classical approach is used for language design, in which the designer creates language only with a perspective of concrete syntax and its semantics – abstract syntax is then completely hidden from designer. In process of language design, when first created part is the model of language – it corresponds to abstract syntax, we should be able to merge different models in order to speak about language composition. Such language models are actually metamodels for sentences written in designed language and in order to compose language models we can use metamodeling composition techniques [3]. Some of the techniques for model composition can be perceived as an abstract syntax composition patterns. Problems in composition of language metamodels include:

- resolution of unique concepts in case of same name in two or more models,
- ways of connecting concepts to each other between different models,
- extending existing language concepts.

Most of these problems could be solved using common modeling techniques, similar to the one known in the object oriented programming. We can have concepts using inheritance and associations for creating links between each other. As for unique concept name it is always possible to use language namespaces, which are common way of differentiating items between name sets – in our case it could be called unique language names.

B. Concrete syntax

Concrete syntax represents way for interaction with language for a language user. It is very important to have as user-friendly representation of a concrete syntax of composed languages as it is possible. Users usually want to write or read languages in known format even if they are composed. Actual methods are aimed on creation of completely new language in case of composition. It is very static way of dealing with a composition problem as it is hard to design a simple change of already composed language, which would automatically reflect in new or evolved version. There are many problems in the area of composition of concrete syntax as textual parsers use context-free grammars for description of language and there are different techniques (LL, LR, etc.) for actual parsing, which all have restrictions. Composition of different languages on level of concrete syntax can be quite difficult process containing many problems, including:

- identifying different tokens consisting of the same written form,
- identifying boundaries between concrete syntaxes of composed languages – how they are connected or embedded,
- parser and scanner techniques restrictions for user-friendly concrete syntax language composition.

These are main problems needed to solve. Solution should allow automatic generation of parser for almost any composed languages which are already formally specified. One of the

possible solutions is to use modular lexical specification language to generate parser like one defined in [4]. It still requires rewriting of language specification to one unified format. Recognition of such problems could lead to specification of concrete syntax language composition patterns which than could be reused for multiple composition cases.

C. Semantics

Next layer of language composition consists of composition of language semantics. Problems concerning composition of language semantics mainly consist of:

- connecting languages with different result sets (computations, generation of code statements, etc.),
- changing semantics of language concept in composition,
- extending semantics of language concept in composition,
- context-aware semantics of concept.

Some of mentioned problems have been addressed in tool LISA [5], some of them are addressed by restrictions. It is almost impossible to automatically compose language semantics which are creating completely different type of result, unless we have translations between such result sets. It is common to find different result sets of semantics defined in already existing domain-specific languages. Another problem is concerning different style of definition of semantics in practical use. Actually, it is out of scope of our work to analyze all problems and possible solutions of composition of language semantics. Possible way, in our case, for at least partial solution of composition of semantics in composed languages is to use semantics written in form of methods in classes in object oriented programming. This way we can use traditional object oriented inheritance, overriding of methods and similar techniques, which can be partially sufficient for testing our composition methodology.

As we have described typical problems in composition of all three language parts, it should be clearer that there is still a big spot for improvement of composition of languages in all areas. Up until now we have always been talking only about textual representation of concrete syntax, but actually any graphical representation of language in computer tends to have some kind of textual serialization form, so it all comes back to processing text.

V. TYPES OF COMPOSITION

Approaches to language composition can be compared to methods used for designing new computer languages. In both cases the result is at least partially new language. We use word *partially* because even methods for designing new languages can sometimes use existing languages. Mernik in his article [6] identifies design patterns for DSL creation using existing languages:

- piggyback (existing language is partially used)
- specialization (existing language is restricted)
- extension (existing language is extended)

Such design patterns are comparable to DSL design patterns described by Spinellis in [7] as long as we concentrate only on the ones connected with reuse of existing languages. The one special design pattern is also pipeline proposed by Spinellis to solve problem of DSL composition.

But his described composition is more of composition in which each language is stand-alone and uses result of previous one as his input. In this case languages connect only in very limited way – they form a pipeline.

For our purposes we have identified some general possibilities of language composition, which are mostly inspired by previously mentioned works in this area. We believe the actual language composition design is not that different from design used for creating new languages. The main difference should be in the ability to maintain specification independence for each of composed language. Languages should be used in the composition without problem of sacrificing later improvements in form of language evolution. Such technique is comparable to usage of programming libraries and APIs.

Identified general language composition patterns are:

- insertion (code is used inside other language)
 - direct/embedding (full parts of code are used)
 - referencing (only identifiers from other code are used)
- specialization (reuse of parts of language)
- extension (full language is used and new parts added)
- combination of all previous patterns

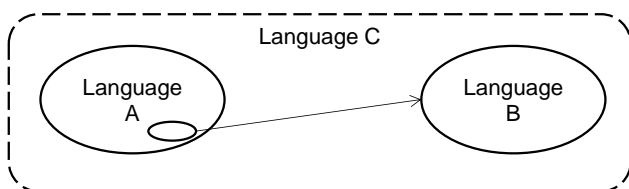


Fig. 1. Composition of languages using insertion pattern

A. Insertion

Most common way of language composition is the insertion of full language code or code fragments into another language. In Fig. 1 we have host *language A* and *language B* which is inserted into host *language A* with purpose to create new composed *language C*. We are not limited to one language, but can have more languages included in one host language. In some cases it could be even hard to specify which language is a host language as languages can interweave in any way. Example of such insertion is usage of SQL database queries inside any general-purpose programming language. In insertion composition pattern we have identified two main subparts: direct and referencing.

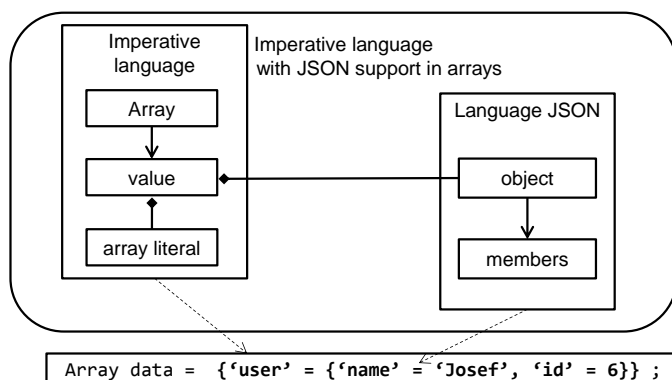


Fig. 2. Composition of languages by direct insertion

1) Direct

Direct insertion composition pattern consists of embedding,

where all code of a *language B* is inserted to a *language A*. We can say that line connecting a *language A* and a *language B* in Fig. 1 is considered as association or full-include relationship. Example of direct insertion composition pattern shown in form of language models is shown in Fig. 2. We have classical imperative language (e.g. Java) with array support and we want to write code for array value in form of JSON syntax, which will be included directly in code of imperative language.

2) Referencing

Specialized way of an insertion composition is the composition by references. In this case there should be unique identifier in referenced language which should be inserted into host language. The actual insertion is then done in way of referencing to identifier by its name instead of using entire code in the place of inserting. In example shown in Fig. 3 we have modeled usual way of using XML files for configuration of computer systems (compare with Spring framework [8]) in which we are referring to class name. This class name is in actual text composed of two parts – package name and actual name of class. Unique identifier for class is then used in XML file. Usually if we change the name of a class or move it to another package such change is not propagated to XML configuration file and configuration can fail. Referencing insertion pattern in language composition can overcome such difficulties by exactly specifying the connection between two languages.

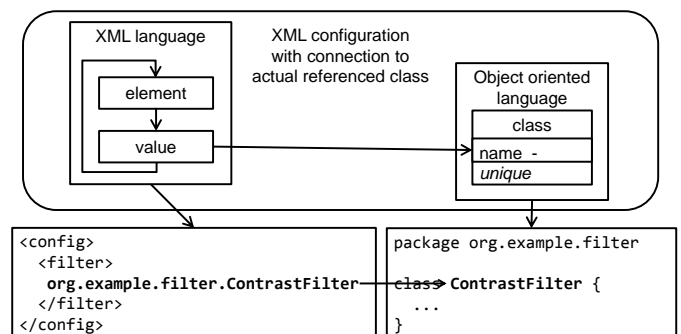


Fig. 3. Composition of languages by referencing insertion

B. Specialization

Specialization composition pattern consist mainly in usage of an existing languages with some restrictions. Those restrictions can specify which language concepts should be used in a new language. Actually this pattern is more of language design pattern than composition pattern, as it does not concern connecting two or more languages. The importance of the pattern lies in reuse with other patterns. It is possible to compose two languages, on which specialization pattern has been used. For example with this pattern we can restrict expression language which provides mathematical operation of division in addition to operations of addition and deduction. In case we do not want to use division in our new language we will simply restrict this original language and leave out the division possibilities.

C. Extension

Opposite to specialization is extension pattern, which can be used to add new possibilities to language by extending it. The original language can be extended by similar method as the inheritance in object oriented programming. This way it

should be possible for original language to stay unchanged. In example shown in Fig. 4 we have expression language with support for addition and multiplication operations. The goal of composition is to make expression language with support for usage of variables in expressions. We extend our original language with language consisting of variables and have new composed language. So the extension composition pattern is actually inheritance transferred to the area of language design.

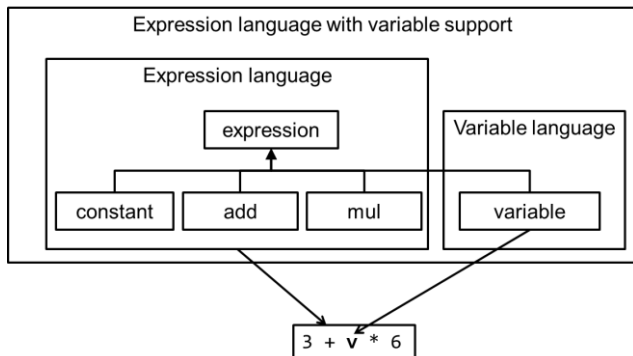


Fig. 4. Composition of languages by extension

D. Combination

Composition of a language is not just simple usage of the one of the previously explained patterns. Usually we need to use more patterns in combination to archive an expected result. Combination of the patterns can also be used to assure immutability of a language used in composition. In case of using the insertion language pattern, we can first use the extension language pattern to add support for the insertion to existing language without modifying it. Such combinations allow language designer to create new language by composing many other in specific way.

VI. COMPOSITION PATTERNS

We aim our research mostly to composition of abstract and concrete syntaxes with a little to no focus on semantics. As it was already mentioned in previous sections there tends to be recurring patterns for language composition which we want to identify in order to simplify the language composition process. We have already presented some of the main general composition patterns. These patterns need to be further enhanced and specialized for exact composition problems described in this paper.

Composition patterns should provide about the same innovation and easier language design as are doing software design patterns in object oriented programming. This way user could be able to reuse known experience in language design as it is done in software design through patterns [9].

VII. CONCLUSION AND FUTURE WORK

Problem of language composition is quite a common now as we have a lot of domain-specific languages being created. There are different tools to support creation of new languages as is MetaEdit+ [10], Xtext [11], YAJCo [12], LISA [5]. Each of these tools is mostly providing support in creation of languages from the scratch or has only minimal support for reuse and composition of existing languages.

In our future work we would like to find more specialized patterns based on our existing language composition patterns and provide processes for reuse of the patterns in language

design process. The patterns need to address composition problems and conflicts and provide solutions at least to the most common problems. As we are aiming our next research on composition of abstract and concrete syntax, we need to solve problems in model composition and composition on lexical and parser levels. There are extensible compiler frameworks like Polyglot [13] and JustAdd [14], which can help in creation of parsers. And for extensible lexing tool generation there is language specification in form of MetaLexer tool, which can than generate JFlex [15] lexer specification from component based specification with support of specifying transitions between components [4].

In the next part of our research it would be best to make actual prototypes for supporting language composition with usage of previously mentioned tools. Such tool will address also problems of composition conflicts and patterns.

ACKNOWLEDGMENT

This work was supported by VEGA Grant No. 1/0305/11 – Co-evolution of the artifacts written in domain-specific languages driven by language evolution.

REFERENCES

- [1] G. L. Steele Jr., "Growing a Language," in *Proceedings of the Conference on Object-oriented programming, Systems, Languages, and Applications*, Vancouver, British Columbia, Canada, 1998.
- [2] J. Greenfield, K. Short, S. Cook, S. Kent and J. Crupi, *Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools*, John Wiley & Sons, 2004.
- [3] M. Emerson and J. Sztiapanovits, "Techniques for metamodel composition," in *The 6th OOPSLA Workshop on Domain-Specific Modeling, OOPSLA 2006*, 2006.
- [4] A. Casey and L. Hendren, "MetaLexer: a modular lexical specification language," in *Proceedings of the tenth international conference on Aspect-oriented software development*, Porto de Galinhas, Brazil, 2011.
- [5] M. Mernik, N. Korbar and V. Žumer, "LISA: a tool for automatic language implementation," *SIGPLAN Not.*, vol. 30, no. 4, pp. 71-79, April 1995.
- [6] 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, December 2005.
- [7] D. Spinellis, "Notable design patterns for domain-specific languages," *Journal of Systems and Software*, vol. 56, no. 1, pp. 91-99, 2001.
- [8] "Spring framework," [Online]. Available: <http://www.springsource.org/spring-framework>. [Accessed 9 3 2012].
- [9] E. Freeman, E. Robson, B. Bates and K. Sierra, *Head First Design Patterns*, O'Reilly Media, 2004.
- [10] S. Kelly and J.-P. Tolvanen, *Domain-Specific Modeling: Enabling Full Code Generation*, Wiley-IEEE Computer Society Press, 2008.
- [11] "Xtext," [Online]. Available: <http://www.eclipse.org/Xtext/>. [Accessed 9 March 2012].
- [12] J. Porubán, 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*, New York, NY, USA, 2010.
- [13] N. Nystrom, M. R. Clarkson and A. C. Myers, "Polyglot: An Extensible Compiler Framework for Java," in *Proc. 12th International Conference on Compiler Construction*, Warsaw, Poland, 2003.
- [14] T. Ekman and G. Hedin, "The JastAdd system - modular extensible compiler construction," *Journal Science of Computer Programming*, vol. 69, no. 1-3, pp. 14-26, December 2007.
- [15] G. Klein, "JFlex: The fast scanner generator for Java," [Online]. Available: <http://jflex.de>. [Accessed 9 3 2012].

Complex program system from ABT view

¹Marián JENČÍK (4th year)

Supervisor: ²Valerie NOVITZKÁ

^{1,2}Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹marian.jencik@student.tuke.sk

Abstract—In this paper we are dealing with construction of the component-based complex program system model which will be used to behavioral description. The basis of the behavioral description of the component-based complex system and definition of the model is *Abstract Behavioral Type (ABT)*. Program component [6] is independent entity, black-box, from which we can compose component-based complex program system and interaction is in this case defined as relation between observable program component input and output through component ports. Complexity of the system is given by number of ABT where individual complex program system we can divided into two ABT types, namely component instances's ABT whose internal structure is unknown and connector's ABT whose internal structure is known and can be defined by user. Behavior of the whole complex system can then be yielded as conjunction of the ABTs.

Keywords—component-based programming, program component, abstract behavioral type, complex program system

I. INTRODUCTION

Component-based software engineering, as part of a software engineering deals with construction of program systems by combining prefabricated components with new programs that provide both glue between the components, and new functionality [5].

In a programming paradigm where instances of components are it's primitive buildings blocks, all decisions about which components will be used and together composed to build new program application must be made from outside what in generally means that a component cannot be allowed to internally decide component will be communicate with another components.

Abstract Behavioral Type (ABT) as model for program components is based on exchanging with it's environment some passive data through it's contact ports and set of the all component's contact port represents interface of the component. Inter-component communication protocol support only two operations, where :

- *sending message* is just write operation and
- *receiving message* is just read operation.

This view of communication among program components is for complex program system modelling very evaluated.

II. COMPLEX PROGRAM SYSTEM

Complex program system P is from the ABT view defined as component-based system, which consists of program component instances pk_i and connectors k_i (*glue code*) between them, which are together modelled as *abstract behavior types (ABTs)*.

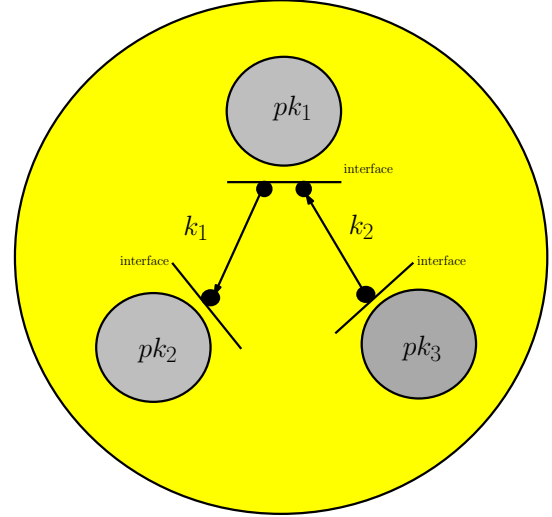


Fig. 1. Complex Program system consists of three program components and two connectors.

Distinction between a component's ABT and a connector's ABT is just that a component's ABT is an atomic ABT whose internal structure is unknown, whereas a connector's ABT is known to be an ABT that is itself composed out of other ABTs

Since *glue code* is between components which connects and through which can components communicate with their environment and it implies that :

- 1) the components must be amenable to external coordination control what implicate to *glue code* must contain suitable inter-communication protocol through which components can interact with their environment and
- 2) the glue code must contain constructs to provide such external coordination what implicate to *glue code* programming language must incorporate a coordination model.

Complex program system P , Fig. 1, can be denoted as :

$$P = \left\{ pk_1 \overset{k_1}{\circ} pk_2 \overset{k_2}{\circ} \dots \overset{k_{n-1}}{\circ} pk_n \right\}. \quad (1)$$

A. Connectors

Definition 2.1: *Connector* can be in general defined as a medium of communication with exactly two ends, and a relation that defines its interaction protocol through these ends. \square

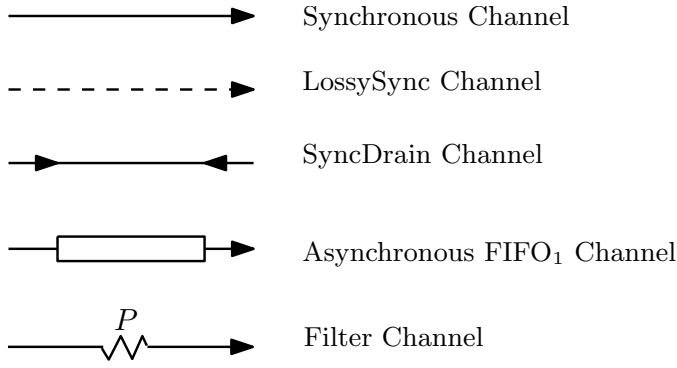


Fig. 2. Some simplest connectors - Channels

We recognize two types of connectors ends:

- *source end* - through this end timed data stream enter into channel;
- *sink end* - through this end timed data stream come out of the channel;

Connector's simplest types are channels (Fig.:2). Channels as a set of simplest *glue code* was firstly defined in REO by Dr. Fährad Arbab [1], [2], [3], [8], [9]. There is no requirement in channels to have defined a source and a sink end. It is perfectly content with a channel that has two sources or two sink, but there have to be defined behavior of the channel through relation which defines interaction protocol of the channel.

Synchronous channel: Behavior of Sync ABT is defined by relation

$$\langle \alpha, a \rangle \text{Sync} \langle \beta, b \rangle \equiv \alpha = \beta \wedge a = b. \quad (2)$$

This notation represents the behavior of any entity in which :

- output data stream is identical to its input data stream ($\alpha = \beta$);
- element on the output is produced at the same time as element on its input is consumed ($a = b$);

LossySync Channel: LossySync is a synchronous channel with a behavior very similar to Synchronous channel except that it is always ready to consume data items written to its source. Behavior of LossySync ABT is defined by relation

$$\langle \alpha, a \rangle \text{LossySync} \langle \beta, b \rangle \equiv \begin{cases} \langle \alpha, a \rangle \text{LossySync} \langle \beta, a(0).b' \rangle & \text{for } a(0) > b(0), \\ \beta(0) = \alpha(0) \wedge \langle \alpha', a' \rangle \text{LossySync} \langle \beta', b' \rangle & \text{for } a(0) = b(0), \\ \langle \alpha', a' \rangle \text{LossySync} \langle \beta, b \rangle & \text{otherwise.} \end{cases} \quad (3)$$

This notation represents the behavior of any entity that :

- all write operations on the source end of a LossySync channel are immediately succeed;
- if there is a pending take on its sink end, then the written data item is transferred;
- otherwise, the write operation succeeds, but the written data item is lost;

SyncDrain Channel: SyncDrain is a synchronous channel and the behavior of SyncDrain ABT is defined by relation

$$\langle \alpha, a \rangle \text{SyncDrain} \langle \beta, b \rangle \equiv a = b. \quad (4)$$

This notation represents the behavior of any entity that :

- produces no output data stream, because it has no sink end;
- every data item written to its source ends is lost;
- write operation on the one of the end remains pending until write on other end is performed, in other world operations on both ends are performed at the same time ($a = b$) and write operations can be blocked if it is necessary, to ensure that they succeed atomically;

Asynchronous channel: The behavior of an asynchronous channel ABT FIFO_k with the bounded capacity of k , for $k > 0$ is defined by relation

$$\langle \alpha, a \rangle \text{FIFO}_k \langle \beta, b \rangle \equiv \alpha = \beta \wedge a < b < a^{(k)}, \quad \text{where } b(i) < a(i+k). \quad (5)$$

The FIFO_k ABT represents the behavior of any entity in which

- output data stream is identical to its input data stream ($\alpha = \beta$);
- every element on its output some time after its respective input element is observed ($a < b$) but before its k^{th} -next input element is observed ($b < a(k)$ which means $b(i) < a(i+k)$, for $i > 0$);

Notation 2.1: Observe that FIFO_1 is indeed a special case of FIFO_k with $k = 1$. FIFO_1 is an asynchronous channel with :

- 1) a source end,
- 2) a sink end and
- 3) bounded buffer with the capacity of one data item.

Its buffer is initially empty. A write operation on its source end succeeds and fills the buffer. If the buffer is non-empty write operation is succeeded on the sink end and removed from buffer. Other input resp. output operations are block and waiting for change the status of the buffer. \square

Filter channel: $\text{FILTER}(P)$ channel behaves like a Sync channel, except that only those data items that match the pattern P can actually pass through it; others are always accepted by its source, but are immediately lost. The behavior of FILTER ABT is defined by relation

$$\langle \alpha, a \rangle \text{FILTER}(P) \langle \beta, b \rangle \equiv \begin{cases} \beta(0) = \alpha(0) \wedge b(0) = \\ = a(0) \wedge \langle \alpha', a' \rangle \text{FILTER}(P) \langle \beta', b' \rangle & \text{if } \alpha(0) \ni P, \\ \langle \alpha', a' \rangle \text{FILTER}(P) \langle \beta, b \rangle & \text{otherwise.} \end{cases} \quad (6)$$

The infix operator $\alpha(0) \ni P$ denotes whether or not the data item $\alpha(0)$ matches with the pattern P and therefore if $\alpha(0)$ passes through or not. When $\alpha(0)$ does not pass through channel, it is lost, and the ABT proceeds with the rest of its timed data streams.

B. Behavior of complex program system

Behavior of complex program system is expressed as conjunction of maximal relations among a set of timed data streams. In other words, complex program system behavior

$B(P)$ is expressed as union of individual ABTs, because abstract behavior type (ABT) is a maximal relation among a set of timed data streams what can be denoted as :

$$B(P) = \{ABT_{pk_1} \cup ABT_{pk_2} \cup \dots ABT_{pk_s}\} \quad (7)$$

or

$$B(P) = \{R_{max}(\langle \alpha_{pk_1}, a_{pk_1} \rangle; \langle \beta_{pk_1}, b_{pk_1} \rangle) \cup R_{max}(\langle \alpha_{pk_s}, a_{pk_s} \rangle; \langle \beta_{pk_s}, b_{pk_s} \rangle)\} \quad (8)$$

III. CONCLUSION

Complexity of the component-based program systems constantly growing up and are more and more used, but if we want to reach expected results, we must be able to model their behavior. Behavior modeling is therefore very important.

Complex program system from ABT view is composed from instances of components and connectors, *glue code* between them which provide interaction, communication among components and both are modelled as abstract behavioral types (ABTs). Program component's ABT is an atomic ABT whose internal structure is unknown, whereas a connector's ABT structure is known and connector's ABT can be itself composed of other connector's ABTs to more complex ABTs.

Inter-component communication protocol of the ABT is based on sending untarget passive messages through component contact ports which only read and write operations are supported. Sending message is just *write operation* and receiving is just *read operation* and set of all component contact ports represents interface of the program component.

Communication between program components, which are independent entities [4] is not as trivial as it in first seems. The blackbox program components do not know anything about each other and are not necessarily designed to work with one another and that means that connector that connects two components and composes pair has to implement an inter-communication protocol which is yielded by relation among observable input and output

$$input_i \xrightarrow{R_k} output_j, \quad (9)$$

where $input_i$ is observable input and $output_j$ is observable output, for $0 < i \leq m$ and $0 < j \leq n$,

or by timed data stream using

$$\langle \alpha_i, a_i \rangle \xrightarrow{R_k} \langle \beta_j, a_j \rangle, \quad (10)$$

where $\langle \alpha_i, a_i \rangle$ is input timed data stream and $\langle \beta_j, b_j \rangle$ is output timed data stream, for $0 < i \leq m$ and $0 < j \leq n$.

ACKNOWLEDGMENT

This work is the result of the project implementation: Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported

by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] ARBAB, F., *Composition by Interaction*. Inaugural Lecture (2005). Leiden Institute of Advanced Computer Science, Faculty of Mathematics & Natural Sciences, Leiden University.
- [2] ARBAB, F., *Abstract Behavior Types: A foundation model for components and their composition*. Science of Computer Programming 55, 2005, pp. 3-52.
- [3] ARBAB, F., *REO: a channel-based coordination model for component composition*. Mathematical Structures in Computer Science 14, 2004, pp. 329-366.
- [4] KRISTENSEN, B. AND MAY, D., *Component Composition and Interaction*. In Proceedings of International Conference on Technology of Object-Oriented Languages and Systems. 1996. TOOLS PACIFIC'96.
- [5] CHEN, X. AND JIFENG, H. AND LIU, Z. AND ZHAN, N., *A Model of Component-Based Programming*. Report No. 350. 2006. UNU-IIST, P.O.Box 3058, Macao.
- [6] SZYBERSKI, C., *Component Software : Beyond Object-Oriented Programming* Addison-Wesley, 1997.
- [7] BAIER, Ch. AND SIRJANI, M. AND ARBAB, F. AND RUTTEN, J., *Modeling Component Connectors in Reo by Constraint Automata* Science of Computer Programming 61, 2006, pp. 75-113.
- [8] CLARKE, D. AND PROENCA, J. AND ARBAB, F. AND LAZOVIK, A., *Deconstruction REO* Electronic Notes in Theoretical Computer Science, Volume 229 (2), 2009, pp. 43-58.
- [9] KOKASH, N. AND ARBAB, F., *Applying REO to service coordination in long-running business transaction*, SAC'09, 2009, pp. 1381-1382.



Component Identification in Company Networks

¹Martin REPKA (3rd year)

Supervisor: ²Ján PARALIČ

^{1,2}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹martin.repka@tuke.sk, ²jan.paralic@tuke.sk

Abstract—This paper deals with social network analysis targeted to decomposition and local structure analysis of organization to organization network. This decomposition is done in order to get view of local structure characteristics in this network and possibilities of correlation or other relationship between compositional data and structural data in Company Networks.

Keywords—Company Network, Network Analysis, Social Networks, Data mining, Knowledge discovery.

I. INTRODUCTION

Company Network (CN) is such a social network, where actors are represented by companies and connections are representing interactions between them. CNs are little bit different from other kinds of social networks, because the actors are not behaving like common actor in social network. Dynamics of CNs is slower and they do not tend to create connections above certain boundaries (local business) because higher trust is needed than in other social networks.

Because of this nature of Company Networks, they are not very dense, but they have local clusters, where density is much higher. This implies that CN will be characterized by significant local structures. If we can detect local clusters containing these structures, then we can simplify and speed up the process of other subsequent network analyses [1].

In next chapters we will take a closer look on CN and we propose a workflow of local structure analysis and possibilities of compositional data utilization in classic network analysis for specific type of social networks. Presented workflow is composed of a weak component clustering analysis, disconnecting of low density clusters and finally identification of local structures.

II. COMPANY NETWORK

A. Motivation

Company Network in Slovakia was built and continuously updated e.g. by ITLIS project¹ [2]. Authors of Itlis aim to save time and costs of their users in finding and processing distributed information. Visualization of this network provides quick look at the direct relationships between people and companies (including companies themselves). CN summarizes all legal relationships between persons and legal persons. All these legal acts are registered in insertion of particular District Court. On the page of ITLIS project¹ user

can find transparent search and visualization tool that lets him/her interactively explore the relationship between the entities. It provides also useful analytical tools to help users find hidden relationships and derived information in the original data. This network could be useful mainly to lawyers, investigative journalists, financial analysts, loan providers and risk managers [2].

B. Origin of Data and Description

Data of Company Network (CN) originate in publicly accessed sources as Business Register of Slovak Republic (ORSR) [3], which was hardly collected, aggregated, analyzed and integrated within project ITLIS.eu into one social network. As a result of these efforts large network was created, featuring complete data model, which integrates structural, compositional and temporal (historical) data [4].

CN is a complex network, similar to two-mode networks It distinguishes three types of actors, as well as three types of connections between actors. Actors represent persons, legal entities (organizations) and agents. All organization actors (OA) must have its own insertion at District Court. Personal actor (PA) can be any person involved in business; all persons involved within boards of companies must have records in insertion of particular company at district court. Agents are actors, which have no insertion in business register, but are mentioned in insertions of other companies. Although it cannot be decided if somebody is a person or legal person in this case, but as a matter of fact it has to be one of them. Usually they are legal persons from other countries.

CN differs from classical two-mode network in following facts. Firstly, it is the presence of agents as third type of actors, but for now we can omit them. So we consider two types of actors: persons (PA) and organizations (OA). Secondly, beside the existence of connections between PA and OA (relationship of affiliation of PA to OA) there is another direct connection between OA to OA, which is not common in classic two-mode networks. We do not consider connections between OA and OA as another type of connections. Specialty of agents is that they have no insertion in Business Register [2].

C. Preprocessing of Company Networks

As an example for demonstration of local structure analysis the organization to organization (O2O) subnet was taken. This O2O subnet is created by direct connections between OAs and PAs and connections between PA and OA were left out. Additional connections could be projected from affiliation

¹ www.itlis.eu

network (PA to OA), but in our case are not considered [5].

After brief analysis we constrained the network by connection quality. Because one relationship “Limited partnership” has significant share (over 90%) of all connections, we chose only to use connections of this quality (O2O LP).

To speed up the whole analysis process, next step is to remove isolated actors and very simple components like isolated dyads. Company network contains a lot of these components and they do not create any special patterns in local structure. After removal of isolated dyads (23% of actors, 32% of connections) we created example sub network (O2O LP*), which we will use in our experiment.

III. ANALYSIS OF LOCAL STRUCTURE

At first we have to examine how much our network is connected. Using analysis for determining connected components we found out, that this network is neither weakly nor strongly connected and there are 2425 weakly connected components [5].

A network is strongly connected if connections have orientation and every actor is reachable from every other actor following the directions of the arcs. On the contrary, a network is weakly connected if its underlying undirected network is connected. A weakly connected network can be thought as an undirected network (we do not consider directions) in which every actor is “reachable” from every other but not necessarily following the directions of the arcs. A strong orientation is an orientation that produces a strongly connected oriented network. A weakly connected component is a maximal subnetwork (same meaning as subgraph) of an oriented network such that for every pair of actors a, b in the subnetwork, there is an undirected path from a to b and a directed path from b to a .

As we observed in our network, there are many local subnetworks disconnected from main component (the largest weak component in network). In other words it means the network is disconnected, which simplifies local structure analysis.

Problem is the largest weakly connected component (11630 actors), which we needed to split into smaller components. For this purpose we can use several actor measures. The simplest and effective method is to remove certain actors. The selection we can do based on actor centrality like in/out degree, total degree, pagerank, betweenness, k -coreness etc.

Betweenness centrality is a measure of a node's centrality in a network equal to the number of shortest paths from all nodes to all others that pass through that node. Betweenness centrality is a more useful measure of the load placed on the given node in the network as well as the node's importance to the network than just connectivity [6].

From logical point of view, the most effective way is to select such actors, which connect dense clusters (local structures) in this component (bridges). For this task removing of actors with high betweenness centrality (disregarding direction of connections) proved to be very successful in disconnecting components in global structure with good enough preservation of local structure.

A. Weak component clustering on O2O LP*

First we have take O2O LP* network and run weak component clustering (WCC), now network is split into clusters (weakly connected components). After size analysis of created clusters (WC Clusters), we classify clusters into small and large cluster groups. We use certain size of cluster as bias, in our case small clusters are those with size less than 150 actors. Small clusters we keep as they are and then we work on splitting of large clusters.

TABLE I.
COMPONENT CLUSTERS IN EXAMPLE SUBNETWORKS

	O2O LP*		LWCC 10750	
	Count	Largest	Count	Largest
Weakly connected	2425	11630	803	87
Strongly connected	25159	10	6988	8

As we mentioned before our O2O LP* network has a very low density ($\sim 4 \cdot 10^{-5}$), same properties have also large disconnected components in this network (from 10^{-4} to $5 \cdot 10^{-3}$). Opposite to large components small ones are quite dense (density from 10^{-2}).

B. Disconnecting large clusters

Using WCC on O2O LP* we extracted one large cluster (11630 actors), which is about to be disconnected. We need to choose suitable actors for removal using their betweenness centrality. After removing of only 5% of actors with highest betweenness centrality it disconnects this large component into a group of small components. This step produced reduced sub network LWCC 10750. Again we use WCC to extract all small clusters.

The size and density of small clusters from disconnected large component is dependent on number of removed top betweenness actors. The reduced sub network LWCC 10750 has similar portfolio of small clusters as O2O LP* “natural” small clusters. Some information about O2OLP* and LWCC 10750 we can find in the Table I.

C. Common Local Structure

After short visualization we can see common structures disconnected or connected in global structures. Most common components are company centered “star” and “line” component. In graph theory, a star S_k is the complete bipartite graph $K_{1,k}$: a tree with one internal node and k leaves. Alternatively, star can be defined as S_k to be the tree of order k with maximum diameter 2, in which case a star of $k > 2$ has $k - 1$ leaves. We will use term “star” in looser definition as component very similar to real star.

Actor centered stars. In most cases stars connect a company with some other companies as satellites. In some cases two or more companies are creating descendant companies and share limited partnership in them. This share implies close cooperation in “central” companies. Remainders of global structure are visible as part of some kind of *lines* or paths. It was usually part of the bridge which was connecting component to the rest of the network. On these lines many

stars may be connected. Line consists mainly from top betweenness actors. New stars are being created from new isolates, or along the lines as result of network growth.

TABLE II.
COMPONENT CLUSTERS IN EXAMPLE SUBNETWORKS

	SN105CL		SN105CL*	
	Count	Largest	Count	Largest
Components	105	-	3	-
Companies	3274	154	268	123

IV. COMPOSITIONAL ATTRIBUTES OF ACTORS WITHIN COMPONENT

Companies in CN have many compositional attributes like financial attributes as basic capital and stock properties or other like list of activities etc. In this paper we will deal with companies' activities which are semi-structured data. This attributes are present in almost all companies.

The companies within component tend to have similar compositional attributes like list of activities. List of activities is semi-structured type of nominal attributes. The activities of companies may slightly vary due to positions within cluster, but the similarity should be statistically significant.

A. Preprocessing of Company activities

For each company list of activities has been retrieved. This list consists of "records", in which each record means a particular activity within field of business. First process is to aggregate activity records to identify the same or very similar records. Aggregation is very important because in many cases the records refer to same activity even if they are not identical (there can be typological mistake, slightly changed word order etc.). From these aggregated activity records words are extracted to represent certain activity of company (attribute is created). In this way we create attribute set, which represents all significant activities from CN.

Another way is to extract set of words as concepts describing certain activity, because the activity could be better described by combination of words, maybe phrase. In our case we used only the words, because concept mining depends of quality of used dictionaries and natural language processing techniques [11].

TABLE III.
PERFORMANCE OF NEURAL NETWORK – CONFUSION MATRIX

	True C1	True C2	True C3	Class precision
Pred. C1	17	3	4	71 %
Pred. C2	4	9	3	56 %
Pred. C3	5	1	8	57 %
Class recall	65 %	69 %	53 %	Accuracy: 63 %

In our experiment we created 4090 numerical attributes (based on activities) for CN. Each of these numerical

attributes represents presence of given activity (range is positive integer, zero for no such activity and other positive number for one or more occurrences of activity in company's compositional data). From the local structure analysis mentioned in Section III we have labeled sub network of 105 components (SN105CL), which consisted of 3274 companies. Label represents that company is connected to certain component. From these 105 components we picked largest 3 components to perform classification based on activity attributes. From these 3 components only suitable companies were chosen due to selection based on missing values in activity attributes. 268 companies out of 308 were suitable for experiment, which is 87 %.

In our experiment we used mentioned set of 268 companies in 3 largest components (C1, C2, C3), so the companies within components are "structurally close" within components and we suppose, that there could be a correlation in compositional data. In Table II. we can see some properties for all components (SN105CL) and our picked 3 largest ones (SN105CL*).

Using RapidMiner2 we tested some classification models on given training set. For performance validation we used (5-fold) cross validation, so the training to testing examples ratio was 0.8. For classification we used several operators using methods e.g. k-NN classification, Bayesian classification (naive Bayes), support vector machines and neural networks.

As most successful model appeared to be *Neural Network* with 1 hidden layer with $\frac{n_a + n_c}{2} + 1$ neurons, where n_a is number of attributes and n_c is number of classes. The best result we achieved can be viewed in Table III.

Operators listed above appeared not to be suitable for this classification task (except neural network), for comparison performance results of Decision Trees were ranging 30-40% (accuracy of classification to the true class).

V. CONCLUSION

In this article we presented proposed workflow of local structure analysis of the Company network. This workflow can be accustomed to any network with similar density and local dense structures. Further we provided experiment, which can suggest, that structural data of the Company Network correlate with compositional data, but this correlation is not so obvious. As we see, the only classification model, which had good enough performance was neural network due to complexity of correlation. In future work we will try the classic form of text mining on companies' activities, and view this attributes as unstructured data, what can solve problem with difficult preprocessing of the activity attributes.

ACKNOWLEDGMENT

This work was partially supported by the Slovak Research and Development Agency under the contract No. APVV-0208-10 (40%) and partially by the Slovak Cultural and Educational Grant Agency of the Ministry of Education of the Slovak Republic under the contract No. 065TUKE-4/2011 (30%). This work is also the result of the project

² <http://rapid-i.com/content/view/181/190/>

implementation Development of Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Programme funded by the ERDF (30%).

REFERENCES

- [1] S. Wasserman, K. Faust, "Social Network Analysis", Cambridge University Press, ISBN 9178-0-521-38707-1, Cambridge, 1994
- [2] P. Kostelník, P. Smatana, Itlis, "Transparent information in context," *Online at www.itlis.eu, 31-Oct-2011*
- [3] Ministry Of Justice Of The Slovak Republic, "Business Register On Internet," *Online at www.orsr.sk, 31-Oct-2011*
- [4] M. Repka, "Analýza určitých typov sociálnych sietí. Pisomná práca k dizertačnej skúške." Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, Košice, Chapter 4, 2011.
- [5] M. Repka, „Local Structure Analysis in Company Network“ In Proceedings SAMI 2012 Herľany, Slovak Republic. ISBN 978-1-4577-0195-5, 2012
- [6] M.E.J. Newman, "Networks: An Introduction," Oxford University Press, Oxford, UK, 2010.
- [7] NWB Team, "Network Workbench Tool. Indiana University, Northeastern University, and University of Michigan 2006, *Online at <http://nwb.slis.indiana.edu>, 31-Oct-2011*
- [8] V. Bagatelj et al., "Generalized Blockmodeling with Pajek", *Metodološki zvezki*, Vol. 1, No. 2, 2004, p. 455-467
- [9] R. Sedgewick, "Algorithms in Java", Third Edition, Part 5 - Graph Algorithms. Addison-Wesley, 2002. ISBN 0-201-31663-3. Section 19.8, pp.205
- [10] G. Tutoky, "Discovery and Exploitation of Knowledge in Collaboration Social Networks" Information Sciences and Technologies. Bulletin of the ACM Slovakia - ISSN 1338-1237. 2011. *Online at <http://acmbulletin.fiiit.stuba.sk/abstracts.html> 31-Oct-2011*
- [11] V. Ročkal, M. Mach, "Concept Mining from Natural Language Texts", In Proceedings SAMI 2012 Herľany, Slovak Republic. ISBN 978-1-4577-0195-5, 2012

A Contribution to Real-Time Network Traffic Evaluation

Adrián PEKÁR (1st year), Marek DUFALA (1st year)
Supervisor: Liberios VOKOROKOS

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

adrian.pekar@tuke.sk, marek.dufala@tuke.sk

Abstract—This paper deals with the development of an Application Programmable Interface (API) for a protocol, that serves for real-time evaluation of network traffic information. This protocol is the Analyzer–Collector Protocol (ACP), which was proposed for direct communication and data transmission between the collecting and analyzing processes of the BasicMeter network metering tool. While the BasicMeter has several components with various application domains, the implementation of direct communication can be a demanding task. So, the development of an API for the ACP would make simpler and more efficient the implementation of real-time network traffic evaluation in the BasicMeter’s modules.

Keywords—IPFIX, ACP, collector, analyzer, real-time data evaluation, network monitoring.

I. INTRODUCTION

One of the most important tasks in any computer network is traffic monitoring. It helps administrators to understand the network and adapt its behavior to any possible situation. By measuring and analyzing network traffic parameters we can secure the network and its users against external or internal attacks [1]; provide smooth functionality of various real-time dependent applications; fulfill the conditions specified in the Service Level Agreement (SLA); deploy Intrusion Detection Systems [2], [3]; or prevent the leakage of sensitive information. All these facts justify the importance of traffic monitoring in the management of today’s modern computer networks and their services.

During the evolution of computer networks many approaches were taken to create a unified standard for monitoring the traffic that flows through the network. One of these approaches is the IP Flow Information Export (IPFIX) [4]. IPFIX defines a format and protocol for the export of information about IP flows. With subsequent analysis of the exported information, traffic monitoring of the network can be achieved.

The BasicMeter metering tool [5] constitutes an alternative solution to the monitoring tools based on the IPFIX protocol. Its main purpose is to provide network traffic parameters measurement and by the ACP protocol real-time data analysis [6]. The architecture of the tool is composed by a set of components, e.g. exporter(s), collector, analyzer(s). While from these components only one collector is needed, from analyzer could be one or more (analyzers may vary in structures and purposes). In this case, the ACP would need to be implemented more than once. While the implementation of this kind of protocol can be in some cases a challenging task, there was a need for an Application Programmable Interface which would

simplify and make more effective the implementation process of the ACP.

In the following sections the BasicMeter tool will be introduced, starting with a brief description of its main components, continuing with a more detailed explanation of the Analyzer–Collector Protocol; up to the description of the proposed solution of the Application Programmable Interface and its implementation details. The last section of this paper draws a conclusion and some future directions.

II. THE BASICMETER METERING TOOL

BasicMeter is a network traffic monitoring tool being developed by the MONICA research group at the Technical University of Košice at the Department of Computers and Informatics. The main purpose of the BasicMeter is to provide network traffic parameters measurement and real-time data analysis [5]. The concept of the metering tool [5] is in conformance to the IPFIX architecture [7]. The architecture of the BasicMeter, as described in Figure 1, consist of the following main components:

- Exporter – is a device that hosts one or more exporting processes. Each exporter is sending flow records containing information about the network traffic to one or more collectors.
- Collector – is a device that hosts a collecting process, which receives flow records from one or more exporting processes in the format of NetFlow v5, v9 [8] or IPFIX protocol [4]. Flow records can be stored in a database for future use or analysis and/or directly sent to one or more analyzing applications by the Analyzer–Collector Protocol (ACP) [6].
- Analyzer – is an analyzing application, which provides a GUI for both the visualization of the information obtained from the database and the management of the architecture’s lower components. However, the analyzer itself is not a subject of the IPFIX architecture [7]. For this reason it was necessary to draft and implement some own network protocols for the data flow [6] and communication between the analyzer and the other components of the metering tool [12].
- Database – is the BasicMeter tool’s data storage.

While the metering tool is based on passive method of traffic measurement [13], the monitoring itself does not require generation of additional network traffic. It makes do with the existing one. Monitoring and analyzing the network traffic characteristics by the BasicMeter tool consist of the following steps:

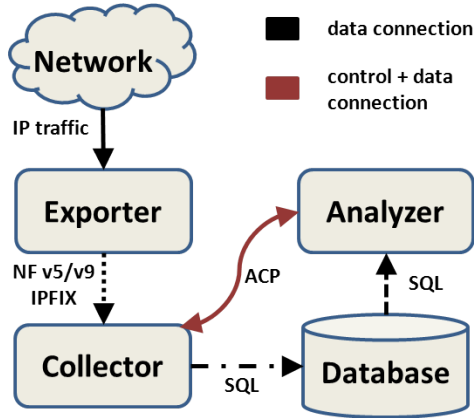


Fig. 1. The architecture of the BasicMeter metering tool

- 1) The information obtained from the captured packets (continuous line in Figure 1) after filtering, classifying, sampling and timestamping are sent by the exporter(s) to the collector (round dot line in Figure 1).
- 2) In the collector, after parsing the currently obtained flow record, the IP flow data are stored in the metering tool's database (dash dot line in Figure 1).
- 3) The analysis over the recorded flow records is performed by the analyzer. By querying the database (dash line in Figure 1), the obtained data are subsequently processed and visualized in a form of flow charts. These charts vary according to the desired type of IP flow characteristics.

III. ANALYZER–COLLECTOR PROTOCOL

In the primary concept of the BasicMeter tool, the measured data should be stored in a database for later analysis. However, this method is absolutely inapt in the case of real-time monitoring. With the growth of data in the database, processing the queries is becoming more and more time-consuming. For this reason there was a need for choosing an other way to gain data about network traffic without the ineffective access to the database.

For this reason was developed the Analyzer – Collector Protocol, which allows the analyzer to gain data about network traffic directly from the collector, without the ineffective access to the database. The collector is permanently inserting the traffic data into a database and if needed, by the means of ACP it is simultaneously sending exactly the same data – or by a customizable filter a part of it – to the analyzer.

ACP is a binary, application layer protocol. The communication is bidirectional and works on the basis of the client-server model, where the client side is represented by the analyzer and the server side is represented by the collector. Over ACP, besides the data, control messages are also transmitted. These messages are dedicated for format check of the received data by the analyzer and also for the control of the communication itself. The types of the control messages are:

- MA – authentication;
- M0 – setting template;
- M1 – setting filter;
- M2 – suspending data transmission;
- M3 – resuming data transmission;
- M4 – setting data transmission type;
- M5 – acknowledging the received data.

The communication is based on sending queries by the analyzer and sending replies with data by the collector. Tables I and II describes the state transitions of the Analyzer – Collector Protocol, where M indicates the type of the control message, TV indicates the truth value of the request and R indicates the particular responses (rejection or acceptance) from the collector.

At the initial state, the collector is awaiting the analyzer's connection at a pre-agreed port (2138 by default). If the connection is established successfully both sides traverse to their S_1 (Start, see Tables I and II) states. Before the bi-directional communication can be opened, the analyzer has to send authentication data (MA) to the collector. On each request the collector responds with an acceptance (TV1) or rejection (TV0). Successful authentication brings both sides to their S_2 (Waiting) states, otherwise the collector terminates the connection.

TABLE I
STATE TRANSITIONS – COLLECTOR SIDE

Old state	Incoming msg. ID	Outgoing msg. ID	New state
S_1	MA_TV0	R0	S_6
	MA_TV1	R1	S_2
S_2	M0_TV0	R0	S_2
	M0_TV1	R1	S_4
S_4	M0_TV0	R0	S_4
	M0_TV1	R1	S_4
	M1_TV0	R10	S_4
	M1_TV1	R11	S_4
	M2_TV0	R20	S_4
	M2_TV1	R21	S_5
S_5	M4_TV0	R40	S_4
	M4_TV1	R41	S_4
S_6	M5	-	S_4
	M3_TV0	R30	S_5
	M3_TV1	R31	S_4
S_6	-	-	-

In S_2 state the analyzer using the template message (M0) can set the desired format of the received data. If the template is accepted, the collector traverses to its S_4 (Transmitting) state and starts with the data transmission in the format of the accepted template.

In the same time, the analyzer traverses to its S_3 (Receiving) state and starts to receive the data from the collector. In this phase the template can be changed anytime. Real data besides control messages are sent/received only in states S_3/S_4 .

By default, the collector exports all traffic information obtained from the exporter(s). However, when necessary, it is also possible to receive only specific information by setting the filter (M1). This optional step is accepted by the collector only in its S_4 state, while only one filter can be set in a time. The filter can be canceled or replaced with an empty filter rule in the collector's S_4 state. Traffic information can be filtered on the basis of source, destination and measuring point's IP addresses; source and destination port; and protocol.

When necessary, the transmission of the data can be suspended (M2) anytime. In such case the collector first sends the remaining data from the currently processed flow record to the analyzer and then both sides traverse to their S_5 (Suspended) states. In general, the collector has no reason to discard the suspension request. The data transmission can be resumed (M3) anytime which brings both sides to their previous states.

TABLE II
STATE TRANSITIONS – ANALYZER SIDE

Old state	Outgoing msg. ID	Incoming msg. ID	New state
S_1	MA_TV0	R0	S_6
	MA_TV1	R1	S_2
S_2	M0_TV0	R0	S_2
	M0_TV1	R1	S_3
S_3	M0_TV0	R0	S_3
	M0_TV1	R1	S_3
	M1_TV0	R10	S_3
	M1_TV1	R11	S_3
	M2_TV0	R20	S_3
	M2_TV1	R21	S_5
	M4_TV0	R40	S_3
S_5	M4_TV1	R41	S_3
	M5	-	S_3
S_5	M3_TV0	R30	S_5
	M3_TV1	R31	S_3
S_6	-	-	-

By default, the collector sends the information elements in N-tuples, which means that during one cycle of data reception the analyzer receives N values. This type of data transmission (M4) can be changed optionally in the collector's S_4 state. Information elements can be also transmitted one-by-one. Data transfer synchronization is ensured by acknowledging the received data (M5).

Unfortunately, receiving data by the ACP also brings a certain delay caused by the processing time in the lower part of the tool. However, this time period compared to the one when obtaining traffic data using a database is significantly less. Thus, from the point of real-time monitoring, gathering traffic data by the ACP is more affordable [12].

IV. APPLICATION PROGRAMMABLE INTERFACE FOR THE ANALYZER–COLLECTOR PROTOCOL (ACPAPI)

As programming language the Java was chosen, which main advantages are in its modularity, stability, perspicuity of the source codes, ability of the documentation auto generation, etc. Except these advantages, Java was chosen also for the reason that the collector was written in Java too. Some of the methods will be used by both parts of the communication, hence of that, the implementation in the same language is more preferable to the implementation in different languages. As mentioned before, the ACPapi has to be usable in the case, when the programmers do not want to waste time with designing and implementing their own classes or methods for the communication based on ACP. For this reason, the whole ACPapi was designed as an easy to understand, implement and use API, besides its development was performed with an emphasis to the annotation of the source codes and its exact documentation.

An API can be written by means of classes or interfaces. According to [14], the most profound feature of Java interfaces is multiple inheritance. With multiple inheritance, only one object is needed to implement an unlimited number of interfaces from an API. In the case of classes, each API class would need a subclass, and if these classes are related to one another, by means of delegation, their instances would need to be joined too [14]. This can significantly increase the amount of occupied memory.

On the basis of this knowledge the ACPapi was written by the means of interfaces. The main advantage is expressed in the performance of the final application using ACPapi. On the contrary, with interfaces, the new methods can be added only by the expansion of the source codes.

Because of the above mentioned dual-usability of some methods, the ACPapi was divided into two packages.

- ACPapi package – contains the API interface itself with declarations of the methods, classes with definitions of the methods and other classes.
- Commonly used package – contains commonly used classes of the analyzer and the collector (classes for working with templates and filters).

As the ACP protocol, the ACPapi defines two types of messages. One for the analyzer and one for the collector. The messages generated by the analyzer describe requests, and the messages generated by the collector contain replies about the acceptance or rejection of the requests. The grammar of the messages are shown below (terminal symbols are highlighted) in Tables III and IV.

TABLE III
GRAMMAR OF THE MESSAGES GENERATED BY THE ACPAPI

Message	→ AuthData Template Filter Pause Resume TransferType DataAccept
AuthData	→ ! md5login md5passwd
Template	→ 0 TemplateData
TemplateData	→ ! InfElementID1..InfElementIDn
Filter	→ 1 FilterData
FilterData	→ ! java serialized Filter object
Pause	→ 2
Resume	→ 3
TransferType	→ 4 TrasferTypeData
TransferTypeData	→ 0 ⇒ One-by-N data transfer (by default) 1 ⇒ One-by-One data transfer
DataAccept	→ 5

Functions defined by Analyzer – Collector Protocol (sending authentication data, templates, filters, etc.) are named upon their functionality. These functions with short description are the following:

- getMd5Digest – method for cipher login and password data,
- connectToCollector – method for establishing the connection with the collector,
- sendTemplate – method for setting the desired format of the received data,
- sendFilter – method for setting filter rules,
- SendPause, sendUnPause – methods for suspending or resuming the data transmission,
- readCollectorAnswers – method for computing the data sent by the collector,
- quit – method for correct termination of the ACP process,
- sendTransferType – method for setting the type of the data transmission.

Main function of the ACPapi is its data providing for almost real-time computation. The word 'almost' points to the fact, that receiving data by ACPapi brings a certain delay with itself, which results from the data computation by the exporter, their transmission over the network, further computation by the collector and sending them by the ACP. Despite of that, from the point of real-time monitoring of network traffic, this delay is absolutely affordable.

TABLE IV
GRAMMAR OF THE MESSAGES RECEIVED BY THE ACPAPI

Message	→ Reply Data
Reply	→ 1 MessageID
MessageID	→ 1 0 10 11 20 21 30 31 40 41
Data	→ 0 DataRecord
DataRecord	→ ! traffic data in the form specified by Template message

Obtaining the data with ACPapi was proposed as a design pattern in which the incoming messages and data are processed in a cycle. The data could be processed by other methods too, but from the point of real-time data processing, the method when the developer is forced to analyze the received data immediately is more favorable.

Since the protocol is defined for communication of both sides, it has to be supported by the collector too. However, a more detailed description of the ACPapi is out of the scope of this paper, but for more information see the ACPapi's website [15].

V. CONCLUSION AND FUTURE RESEARCH

The basic motivation of the reported research was given by the importance of accurate and real-time evaluation of traffic characteristics. Current researches and developments on the field of network monitoring are leading to the reduction of the data evaluation and analysis time of the monitoring tools.

In the previous sections the main concept of the BasicMeter metering tool and its ACPapi for real-time monitoring was presented. The ACPapi fulfilled its expected results (see some of the other works of MONICA [15]). Thanks to the proposed design pattern and some improvements to the ACP (setting transfer type or acknowledging the received data) the ACPapi gives reliable, easy to use and implement application interface for the developers of the BasicMeter measuring tool.

Future work should be aimed at the ACP protocol and the ACPapi too. In the current version of the ACP, the templates are sent as arrays containing the identifiers of the information elements. However, in the future we plan to replace the arrays with a common binary description of these elements. In case of ACPapi, future work should be aimed at the improvement of the actual design pattern.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] L. Vokorokos, N. Ádám, A. Baláz, "Application Of Intrusion Detection Systems In Distributed Computer Systems And Dynamic Networks," *Computer Science and Technology Research Survey, CST 2008*, Košice, Elfa, pp. 19–24, ISBN 978-80-8086-100-1, 2008.
- [2] L. Vokorokos, A. Kleinová, O. Látka, "Network Security on the Intrusion Detection System Level," *The 10th IEEE International Conference on Intelligent Engineering Systems, INES 2006*, London, United Kingdom, pp. 270–275, ISBN 1-4244-9708-8, 2006.
- [3] M. Tomášek, M. Čajkovský, B. Madoš, "Intrusion Detection System Based on System Behavior," *10th IEEE Jubilee International Symposium on Applied Machine Intelligence and Informatics, SAMI 2012*, Slovakia, pp. 271–275, ISBN 978-1-4577-0195-5, 2012.
- [4] B. Claise, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information," RFC 5101, 2008.
- [5] F. Jakab, Ľ. Koščo, M. Potocký, J. Gierl, "Contribution to QoS Parameters Measurement: The BasicMeter Project," *Conference Proceedings of the 4th International Conference on Emerging e-learning Technologies and Applications, ICETA 2005*, pp. 371–377, ISBN 8080860166, 2005.
- [6] J. Gierl, R. Jakab, Ľ. Koščo, "Communication Protocol in Computer Network Performance Parameters Measurement" *4th International Information and Telecommunication Technologies Symposium, I2TS 2005*, Florianopolis, Brazil, pp. 161–162, ISBN 85-89264-05-X, 2005.
- [7] G. Sadasivan, N. Brownlee, B. Claise, J. Quittek, "Architecture for IP Flow Information Export," RFC 5470, 2009.
- [8] B. Claise, "Cisco Systems NetFlow Services Export Version 9," RFC 3954, 2004.
- [9] A. Baláz, J. Trelová, M. Kostráb, "Architecture of Distributed Intrusion Detection System Based on Anomalies," *14th International Conference on Intelligent Engineering Systems*, Las Palmas of Grand Canaria, Spain, pp. 79–83, ISBN 978-1-4244-7651-0, 2010.
- [10] J. Quittek, T. Zseby, B. Claise, S. Zander, "Requirements for IP Flow Information Export (IPFIX)," RFC 3917, 2004.
- [11] T. Zseby, E. Boschi, N. Brownlee, B. Claise, "IP Flow Information Export (IPFIX) Applicability," RFC 5472, 2009.
- [12] F. Jakab, J. Gierl, R. Jakab, M. Kaščák, "Improving Efficiency and Manageability in IPFIX Network Monitoring Platform" *6th International Network Conference, INC 2006*, Plymouth, United Kingdom, pp. 81–88, ISBN 1-84102-157-1, 2006.
- [13] J. Sučík, F. Jakab, "Measurement and Evaluation of Quality of Service Parameters in Computer Networks," *PhD. Conference and ŠVOS FEI TU of Košice*, Košice, pp. 119–120, ISBN 809686663X, 2003.
- [14] J. Tulach, "Practical API Design: Confessions of a Java Framework Architect," Apress, ISBN 978-1430209737, 2008.
- [15] The MONICA Research Group, "Application Programmable Interface for serving the Analyzer–Collector Protocol (ACPapi)," [Online]. Available: <http://wiki.cnl.sk/Monica/ACPapi>

Creation of Depth Map for 3D Display

Peter IVANČÁK (1st year), František HROZEK (3th year)
Supervisor: Branislav SOBOTA

Dept. of Computers and Informatics, FEEI TU of Košice, Slovak Republic

peter.ivancak@tuke.sk, frantisek.hrozek@tuke.sk, branislav.sobota@tuke.sk

Abstract— This paper aims on problematic of content creation for 3D displays and is divided into two parts. First part focuses on various displaying technologies and 3D displays, especially stereoscopic displays passive, active and auto-stereoscopic. Second part presents autostereoscopic 3D display Philips WOWvx and its displaying method (2D-plus-depth). Second part also presents application that calculates depth map from stereoscopic image. This application use Block Matching, Graph Cut and Semi-Global Block Matching algorithm for depth map calculation and was developed at DCI FEEI TU of Košice.

Keywords— 3D displays, 3D displaying technologies, 2D-plus-depth, depth map calculation.

I. INTRODUCTION

3D displays have become a common part of everyday life. Compared to 2D display, 3D provides additional depth information and represents objects in a more natural and accurate way. There are several technologies that can be used for 3D display: stereoscopy, volumetric visualization and holography [1]. Research at DCI FEEI TU of Košice is focused on stereoscopic technologies.

Stereoscopy is based on the way how human brain perceives the surrounding objects. When a person looks at an object, it is seen by each eye from a slightly different angle. Human brain processes the information and enables stereoscopic vision. A stereoscopic image (or a video) is a composition of left and right images captured by the two cameras. The image is then presented to the viewer in such manner that each eye can see its image (left eye see left image and right eye see right image). In this way human brain interprets a 2D image as a 3D scene.

This paper aims on autostereoscopic displays which allow users to see 3D image without glasses or any headgear. Paper also presents application that was developed at DCI FEEI TU of Košice. This application calculates depth map from stereoscopic image or video using three algorithms (Block Matching, Semi- Global Block Matching and Graph Cut). This depth map is used for 3D image creation in Philips WOWvx.

II. 3D DISPLAYING TECHNOLOGIES

3D displays use several technologies to create 3D image. Each technology has its advantages and disadvantages. There are several types of 3D displays [2]:

- Holographic displays
- Volumetric displays (swept-volume and static-volume displays)
- Stereoscopic displays
 - Passive (anaglyph, polarized stereo, INFITEC - Interference filter technology)
 - Active
 - Autostereoscopic (parallax barrier, lenticular lens)

A. Anaglyph

It has become quite a popular stereoscopic method, since it only requires usage of very cheap red-cyan glasses. The anaglyph is a composition of a red channel from the left image and green and blue channels (together creating a “cyan channel”) from the right image. When looking at the anaglyph through red-cyan glasses, the left eye can only see the red part of the image and the right eye the rest. Anaglyph image is only one stereoscopic method that can be printed on paper.



Fig. 1. Example of anaglyph

B. Active stereoscopy

Active stereoscopy is now probably the most common method used for 3D TVs. It has high picture quality. Frame rate of 3D displays which use this technology is doubled (120 Hz – 60 Hz for each eye). User need for viewing special electronic glasses that are synchronized with the remote broadcast source and alternately shows images for the left and for the right eye.

C. Autostereoscopy

Autostereoscopy is method of displaying stereoscopic images without the use of special headgear or glasses on the part of the viewer. There are two autostereoscopic methods: *active* (special camera track user's eyes position) and *passive* (no special hardware is needed). Passive autostereoscopy use two most common technologies for displaying of 3D image: *parallax barrier* (Fig. 2 up) and *lenticular lens* (Fig. 2 down).

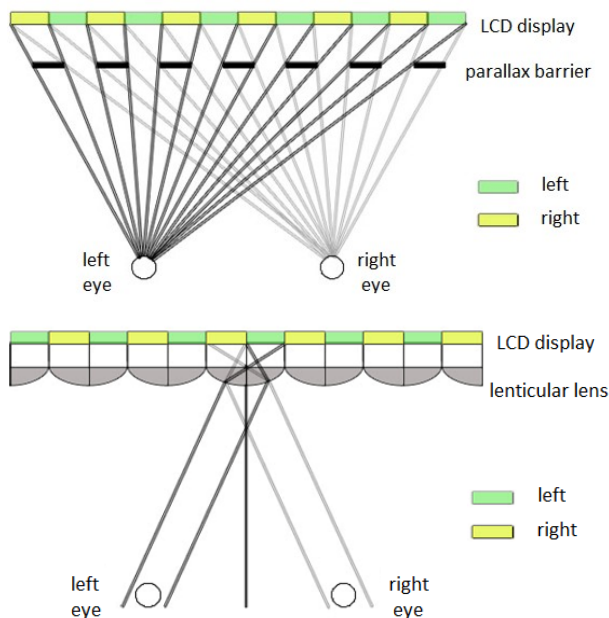


Fig. 2. Passive autostereoscopic displaying technologies: (up) parallax barrier, (down) lenticular lens

III. 2D-PLUS-DEPTH TECHNOLOGY

Another passive autostereoscopic technology use 3D display Philips WOWvx. 3D image is created using 2D-plus-depth technology. This technology creates 3D image using 2D image and depth map. Example of input image for Philips WOWvx is shown Fig. 3. More details about this 3D display and 2D-plus-depth method at manufacturer's webpage [3].

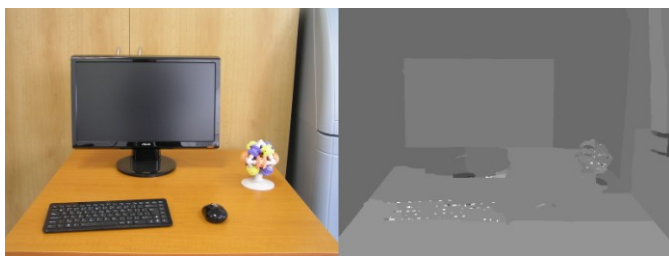


Fig. 3. Input example for Philips WOWvx: 2D image (left half) and depth map (right half).

A. Application for depth map calculation

This application was developed at DCI FEEI TU of Košice. Input for this application is static image (left and right image or stereo image) or dynamic image (stereo video).

For depth map calculation application uses these three algorithms:

- *Block Matching (BM)* – is a way of location matching blocks in a sequence of digital video frames for the purposes of motion estimation.
- *Semi-Global Block Matching (SGBM)* – is more accurate than standard BM and with almost the same calculation speed as standard BM. This algorithm tries to find the best possible alignment between two images.
- *Graph Cut (GC)* – normal stereo matching algorithms try to match a pixel in the left image to a pixel in the right image based on some individual property like color. Although this is fast and reasonably accurate process it does not deal with interlinear consistency. Graph Cut algorithm incorporates interlinear consistency and takes in account the interpixel properties like adjacency. This results in a much more accurate disparity depth map.

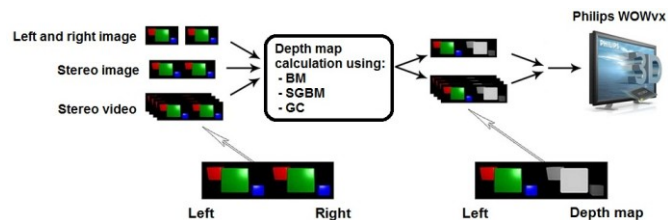


Fig. 4. Creation process of depth map for Philips WOWvx

Two types of tests were conducted on developed application – calculation time and depth map quality test. Hardware configuration of computer where application was tested:

- CPU: Intel Core i5 3,20 GHz
- HDD: 500 GB 7200 RPM 16MB Cache
- RAM: 4 GB DDR 3 1066 MHz
- GPU: NVIDIA GeForce 275 GTX

Calculation time test

Fastest algorithm for dept map calculation is BM, second one is SGBM and the last one is GC. Dependency between calculation time and image resolution for each one algorithm are shown in Fig. 5, 6, 7.

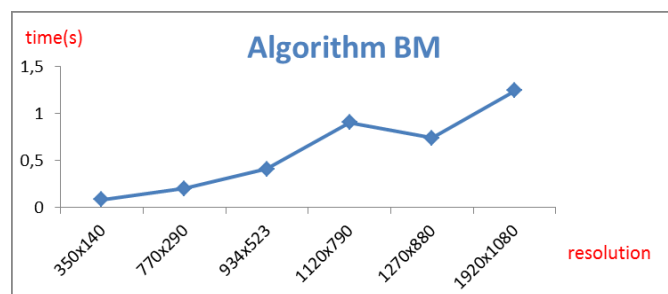


Fig. 5. Dependency between calculation time and image resolution (BM)

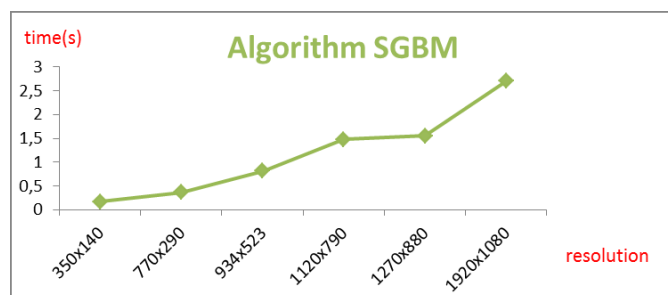


Fig. 6. Dependency between calculation time and image resolution (SGBM)

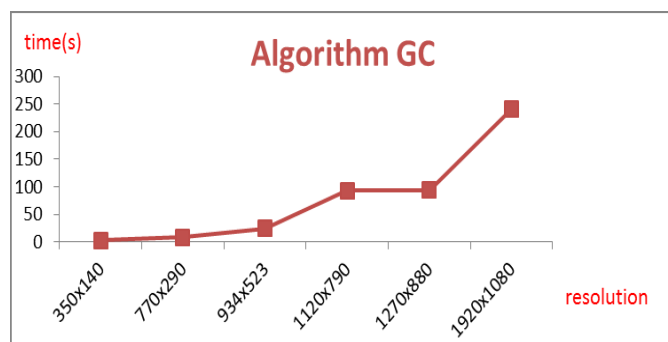


Fig. 7. Dependency between calculation time and image resolution (GC)

Depth map calculation time is dependent from these six factors: image color, image resolution, number of objects in image, relative position of objects, objects distance and image background. For example calculation time of first image (Fig. 8 up; resolution 640×480, more objects and complex background) was 99 seconds. Compare to this, second image calculation time (Fig. 8 down; resolution 800×600, fewer objects and simple scene) was only 40 seconds.



Fig. 8. Two images with different depth map calculation times (up 99 seconds and down 40 seconds)

Depth map quality test

The best depth map quality has GC, second one is SGBM and the last one is BM. Fig. 9 shows 2D image and depth maps calculated by individual algorithms.

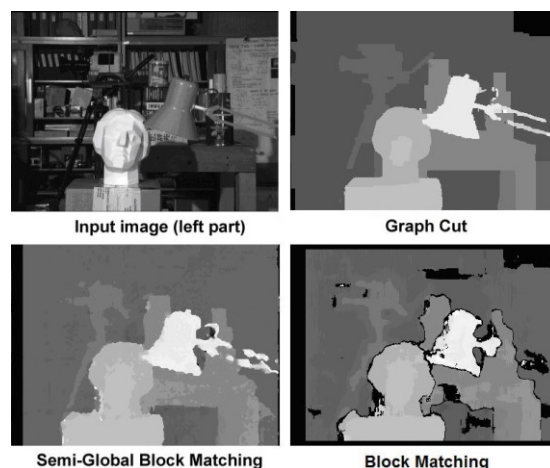


Fig. 9. Source image and calculated depth map for individual algorithms tested in quality test

IV. CONCLUSION

In this paper was presented application developed at DCI FEEI TU of Košice which calculates depth map from stereo image (video) using three algorithms: Block Matching, Semi-Global Block Matching and Graph Cut. Each one algorithm has its advantages and disadvantages. As can be seen from conducted tests (Fig. 5-9), BM algorithm has fastest calculation time but worst quality of calculated depth map. GSBM offers better quality of depth map with also good calculation time. GC has longest calculation time but calculated depth map is very good.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] Sobota, B., Szabó Cs., Perháč, J., Ádám N.: 3D Visualization for City Information System, Proceedings of International Conference on Applied Electrical Engineering and Informatics AEI2008, Athens, Greece, 8.-11.9.2008, Košice, FEEI TU Košice, 1, ISBN 978-80-553-0066-5, 9-13.
- [2] Banz, C.; Hesselbarth, S.; Flatt, H.; Blume, H.; Pirsch, P.: "Real-time stereo vision system using semi-global matching disparity estimation: Architecture and FPGA-implementation," *Embedded Computer Systems (SAMOS), 2010 International Conference on*, vol., no., pp.93-101, 19-22 July 2010 doi: 10.1109/ICSAMOS.2010.5642077.
- [3] Philips WOWvx homepage, <http://www.business-sites.philips.com/3dsolutions/home/index.page>.
- [4] Vicente, S.; Kolmogorov, V.; Rother, C.: "Graph cut based image segmentation with connectivity priors," *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, vol., no., pp.1-8, 23-28 June 2008 doi: 10.1109/CVPR.2008.4587440.
- [5] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger, "Depth Map Creation and Image Based Rendering for Advanced 3DTV Services Providing Interoperability and Scalability", *Signal Processing: Image Communication. Special Issue on 3DTV*, February 2007.
- [6] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16-22, 2000.
- [7] de Haan, G.; Biezen, P.W.A.C.; Huijgen, H.; Ojo, O.A.; "True-motion estimation with 3-D recursive search block matching," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol.3, no.5, pp.368-379, 388, Oct 1993

Decision-making framework for adoption of cloud computing

¹Adela TUŠANOVÁ (2st year)

¹Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹adela.tusanova@tuke.sk

Abstract— the popularity of cloud computing and its potential is leading to a growing number of organizations considering its implementation. Investing in this new technology and its adoption may be the most important decision organizations will face. Unfortunately, cloud computing adoption is not as easy as other information technology adoption. Therefore careful consideration of financial and non-financial criteria should be included in the analysis. In this paper I introduce conceptual decision-making framework, based on multi-criteria, which takes readiness for cloud adoption into the account.

Keywords— cloud computing, cloud adoption, decision-making, multi-criteria, framework

I. INTRODUCTION

Cloud computing adoption represents for IT managers decision-making, while they are always looking for the best solution. The aim of the transition to the cloud may be for example lower costs, increased effectiveness of business processes, improvement of system availability and reliability, ability to scale IT resources etc. These motivation factors are most often leading to cloud computing adoption in companies.

The question, whether adopt cloud or not is one of the first questions that IT managers need to ask. It is also important to decide on the correct workload to be transferred to the cloud and select the best alternative with respect to conditions of the organization. Since this is a complex process of decision making, the basic framework for support of this process is more than necessary. Therefore, I present a framework to facilitate the process of decision-making with respect to cloud adoption in this paper.

II. RELATED WORK

There are several approaches which support decision making during cloud computing adoption. Some approaches are general, providing only a kind of overview of all possible alternatives, or sequence of steps to be completed in the transition to the cloud. Other approaches take into account only costs, and do not take into account any benefits that arise from the transition. Sophisticated solutions see cloud computing adoption as more complex problem and thus they try to solve decision problem by using a variety of multi-criteria decision making methods.

One of the basic approaches by IBM [1] provides a general approach to this issue - explains all the alternative options

provided by cloud computing (SaaS, PaaS, IaaS), but also private, public or hybrid cloud). However, it does not provide an answer if company is ready for adoption or which solution is the best.

Microsoft's framework [2] provides an analysis of the application into two segments and then assess whether the application is or is not appropriate for the transition to cloud computing. It focuses only on SaaS solutions and does not take into account the transition costs and potential benefits that cloud solutions deliver.

Researchers from universities in Greece [3] focused on the analysis of the costs incurred in developing software solutions. By analyzing and comparing the resulting costs they answer the question whether it is worthwhile to develop applications in the cloud or on-premise. This approach is based solely on costs appraisal; it does not take into account other benefits which arise from the cloud.

Other approaches bring scientists from FZI in Germany [4] as well as scientists from the Institute of Technology in India [5]. They tried to answer the question adopt or do not adopt cloud services, focusing on infrastructure as a service. The first approach is based solely on costs appraisal and does not take into account other benefits; the second approach is based on decision-making and takes other criteria and their weights into account.

The most comprehensive approaches offer scientists from IBM Research [6] and Michael Menzel et al. [7]. Both solutions are based on multi-criteria decision methods. The first approach tries to answer the question, which service model and deployment model of cloud computing to implement and which provider to choose (this option is only outlined). The second method is focusing on infrastructure as a service. It does not reflect only the costs but also positive aspects of cloud computing adoption. The outcome of both approaches is most appropriate alternative based on chosen multi-criteria method. These approaches, however, do not analyze organization's suitability for particular cloud solution in the context of entire organization and do not answer the question whether the organization is ready to adopt or not. Also, costs and other economic indicators are listed in the model only as a criterion, which are estimated in advance, but do not address their actual calculation.

III. THE FRAMEWORK

Adoption of cloud computing is affected by many factors and each organization is unique. In practice, when an individual analysis of the whole organization is made, there are several analysis performed - technical analysis (analysis of IT infrastructure, processes, applications ...), economic analysis (costs, incomes, revenues ...) or strategic analysis. To carry out these analyzes specialists are needed, whose analysis is expensive and takes several weeks. Therefore, I strive to create a model that would help IT managers in decision making.

The purpose of our framework is to help IT managers to determine best alternative of cloud services for their organization. To make as good decision as possible, we need to make decision in the context of specific workload. Without specific workload defined, decision framework would be too abstract. Thus, for our framework I assume that the IT manager has ideas which workload could be transferred to the cloud. The framework leads IT managers step by step from workload definition to choose the most appropriate alternative. The goal is to help IT managers to evaluate if the workload is appropriate for the adoption to cloud and which alternative is the best. One of the most important factors which influence decision making is costs; therefore a key step of the process is the calculation of Total Cost of Ownership (TCO).

The framework itself consists of the 6 steps (as Fig. 1 shows).

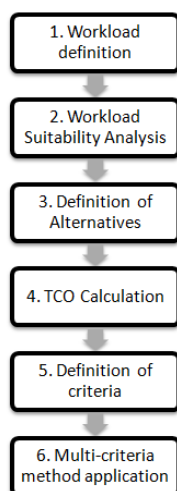


Fig. 1. Cloud computing adaptation framework

A. Workload definition

We understand workload as the kind of work that an organization needs to accomplish. In other words, workloads represent collections of key IT system components and the relationships among them. These components can include web servers and application servers, databases and behavioral policies such as availability, security and performance. Each workload has characteristics that make it run most efficiently on certain types of hardware and software. There are also other requirements, for example some workloads demand fast transactions, like ATMs, while others, like predictive analytics, require intense calculations [[8]], [9]. For the purposes of our model, we have to create workload taxonomy according to a cloud service model. There are three main types

of cloud workload - Software as a service workload type (includes applications, e.g. e-mail, ERP applications), platform as a service workload type (includes programming languages, tools and platforms to develop, deploy and manage applications) and infrastructure as a service workload type (includes processing, storage, networks and other computing resources). Defined type of workload (IaaS, PaaS, SaaS) is important input to the second step of our process - Workload suitability Analysis.

B. Workload suitability Analysis

The purpose of this step is to ask IT manager a set of questions, which will help to evaluate suitability or unsuitability of particular workload for cloud computing. The analysis consists of a simple checklist of questions to provide a rapid assessment of the suitability of a particular workload for a specific organization. The appropriateness of adopting a given workload in a particular organization is influenced by several factors. Therefore, there are sets of questions concerning technical, legal and economic areas. For each type of workload (based on cloud services models) is prepared set of questions under specific cloud services models. After assessment of all answers, if the workload is suitable for cloud, IT manager can continue with next step of our decision-making process – definition of alternatives.

C. Definition of Alternatives

When cloud computing workload is under consideration, there is almost never just one solution. Thus, there is a problem which solution to choose. The problem may be defined as finding the best alternative for a decision maker (IT manager). Always, at least two options should be defined for particular workload. If on-premise solutions in organization already exist, it can be very interesting compare this solution with possible cloud alternative(s) and make right decision (e.g. no action vs. cloud adoption). Each alternative is characterized by a set of criteria, defined in 5th step of our decision-making process.

D. TCO Calculation

One of the most important factors affecting the decision on the adoption of any IT solution is cost. To calculate the cost of each alternative in our framework, the Total Cost of Ownership (TCO) method is used.

TCO expresses the total cost of implementation and operation of IT (e.g. direct costs on hardware and software, operation and maintenance costs, administration costs, users training, costs of inactivity due to system patches, updates, etc.). In the context of cloud computing we are calculating TCO for each cloud service model differently. Calculation of TCO of infrastructure as a service was introduced by [4], calculation of application development in cloud was introduced by [3] **Chyba! Nenašiel sa žiaden zdroj odkazov.** and calculation of TCO of software as a service was introduced by [10], [11]. Complex cloud TCO taxonomy for calculation is still missing; therefore I plan to focus our research also in this area.

E. Definition of criteria

Identify criteria means provide additional information on which alternatives are considered. During the criteria design it

is important to keep in mind that a good criterion is only such one that has the sense and benefit in context of particular choice. We have to know what we require and we have to be able to formulate criteria in such a way that we will be able to make clear selection.

In general, there are four types of criteria: maximization, minimization, which is also always a combination of quantitative and qualitative criteria. The maximization criterion says if it is preferable if its value increases (higher value is better than less). An example of such criterion is income. The minimization criterion prefers if its value falls (lower value is better as above). An example of such criterion is costs. Quantitative criterion is a criterion which is clearly measurable and it can be attributed to the value expressed in a number. An example of such criteria is price, income, etc. Quality criterion is not directly measurable; therefore it is important to establish how it will be measured. Appropriate methods appear to be scoring or grading. Criteria are always a combination of these groups, therefore four possible combinations arising: quantitative maximization criterion, quantitative minimization criterion, qualitative maximizes criterion and qualitative minimization criterion. Depending on the multi-criteria method it may be necessary to convert qualitative criteria to quantitative.

F. Multi-criteria method application

In decision-making process there are several different methods, which are essentially based on the same principle, and it is assessing several alternatives to address the problem by selected criteria (and possibly determination of the order of alternatives). These methods vary according to how they determine the weight of each criterion and how they determine numeric degree, which is assigned to all alternatives for each criterion.

For purposes of this framework, I propose to apply Decision Matrix Method (DMM) [12]. This method, also known as Pugh method or Pugh decision matrix is considered as the basic method of multi-criteria decision making. It solves problem with more than one alternatives of the solution. If needed, advanced multi-criteria method like Analytic Hierarchy Process or Analytic Network Process can be used.

IV. CONCLUSION AND FUTURE WORK

Cloud computing presents for organizations opportunities such as costs reduction, risk reduction and increase of competitive advantages. Adopting cloud computing is a complex issue that requires comprehensive analysis of all alternatives.

In this paper, I introduced a generic decision-making framework for cloud adoption. It defines step-by-step decision-making process from workload definition to selection of appropriate alternative. As I know, this is the first cloud decision-making framework, which takes into account also workload suitability and TCO. It is designed for all cloud services types (SaaS, PaaS, and IaaS).

Fields of future work include research of suitability in cooperation with experts. Also, comprehensive analysis of costs and construction of TCO taxonomy for all categories of workloads is needed. I see high potential of this framework in

particular for small and medium enterprises as well as organizations of all types which are using various IT services in their work. Therefore it is planned to be implemented as web-based decision support system.

ACKNOWLEDGMENT

The work presented in this paper was partially supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic under grant No. 1/1147/12 (50%). The work is also the result of project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (50%).

REFERENCES

- [1] IBM: Defining a Framework for Cloud Adoption, http://www.idgconnect.com/view_abstract/6194/defining-framework-cloud-adoption
- [2] Microsoft Slovakia: Microsoft & Cloud, <http://itevent.sk/cloud/webcast/microsoft-cloud>
- [3] Stamatiou Bibi, Dimitrios Katsaros, Panayiotis Bozaris, "Application Development: Fly to the Clouds or Stay In-house?," Enabling Technologies, IEEE International Workshops on, pp. 60-65, 2010 19th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises, 2010
- [4] Kleems M., Nimis J., Tai S.: Do Clouds Compute? A Framework for Estimating the Value of Cloud Computing, Lecture Notes in Business Information Processing, 2009, Volume 22, Part 4, 110-123, DOI: 10.1007/978-3-642-01256-3_10
- [5] Subhas Chandra Misra, Arka Mondal, Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding Return on Investment, Mathematical and Computer Modelling, Volume 53, Issues 3-4, Telecommunications Software Engineering: Emerging Methods, Models and Tools, February 2011, Pages 504-521, ISSN 0895-7177, DOI: 10.1016/j.mcm.2010.03.037.
- [6] Saripalli, P.; Pingali, G.; , "MADMAC: Multiple Attribute Decision Methodology for Adoption of Clouds," Cloud Computing (CLOUD), 2011 IEEE International Conference on Cloud Computing, pp. 316-323, 4-9 July 2011 doi: 10.1109/CLOUD.2011.61 URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6008725&isnumber=6008659>
- [7] M. Menzel, M. Schönherr, J. Nimis, St. Tai: (MC²): A Generic Decision-Making Framework and its Application to Cloud Computing, International Conference on Cloud Computing and Virtualization (CCV 2010), Singapore, May 2010
- [8] IBM: Dispelling the vapor around cloud computing, <http://public.dhe.ibm.com/common/ssi/ecm/en/ciw03062usen/CIW03062USEN.PDF>
- [9] IBM: Get more out of cloud with a structured workload analysis, <http://public.dhe.ibm.com/common/ssi/ecm/en/iaw03006usen/IAW03006USEN.PDF>
- [10] Aggarwal Sanjeev, McCabe Laurie: The TCO Advantages of SaaS-Based Budgeting, Forecasting & Reporting, An Analysis of the Four-Year Total Cost of Ownership (TCO) for SaaS and On-Premise Performance Management Solutions, http://www.adaptiveplanning.com/docs/Hurwitz_TCO_of_SaaS_CPM_Solutions.pdf
- [11] Software-as-a-Service Executive Council: Software-as-a-Service A Comprehensive Look at the Total Cost of Ownership of Software Applications, <http://www.winnow.com/saas.pdf>
- [12] S. Pugh: Concept selection: a method that works. In: Hubka, V. (ed.), Review of design methodology. Proceedings international conference on engineering design, March 1981, Rome. Zürich: Heurista, 1981, blz. 497 – 506

Denial of Service Intrusion Detection System

Anton BALÁŽ

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

anton.balaz@tuke.sk

Abstract—Denial of Service attacks are significant threat to computers and computer networks. Actually there are lots of variations of DoS. Are distinguished by the methods and forms of attack. Largest attacks are a major threat for large companies and still there is few effective defence against attacks. The main concern of this work is to design methods to detect certain types of DoS attacks in computer networks.

Keywords—DoS, IPFIX, network, attack, intrusion, IDS

I. INTRODUCTION

Denial of Service is a type of attack, where attacker tries to disable the target device from service. There are several methods to realize it, but the essence of attacks is to overload target with false requests for service that causes the justified requests can't be served. Attack on certain types of services with imperfect debugging can cause a fall of service or crash the whole system. DoS attack is usually considered non-destructive and affects the target system only for the duration of the attack. On the other hand, there are DoS attacks whose consequences of which persist even after the attack. A typical example is a DoS attack, which causes a lack of space for storage devices. What is usually caused by a disproportionate increase in the size of log files.

IP Flow Information Export (IPFIX) is working group to specify the information model to describe the IP flows and defines the IPFIX protocol that serves to transmit information of the flows from exporters to collectors. IPFIX system can provide information that can help to detect DoS attacks. [1], [2]

BasicMeter is a platform that implements the IPFIX system. Platform develops at the Department of Computers and Informatics TU Košice. [3]

The aim of this work is to analyze the traces of certain types of Denial of Service attacks in the image of network traffic. On these tracks implement intrusion detection system. The system should be able detect several types of DoS attacks: SYN flood, UDP flood and HTTP flood. Image of network traffic will be obtained from the IPFIX system.

II. DESIGNED ARCHITECTURE

The detection system was designed as application module for measuring platform BasicMeter [3]. The proposed system obtains data of network operation from the database where the collector stores the data. User interface detection system is a simple web interface that is connected with the core system through its own database (Fig. 1).

The core of the detection system is an application written in C++ language. Takes the form of console applications with extra work as a system service. It consists of a group of

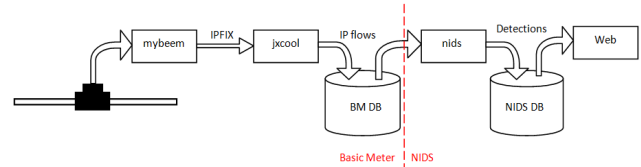


Fig. 1: Architecture of the detection system

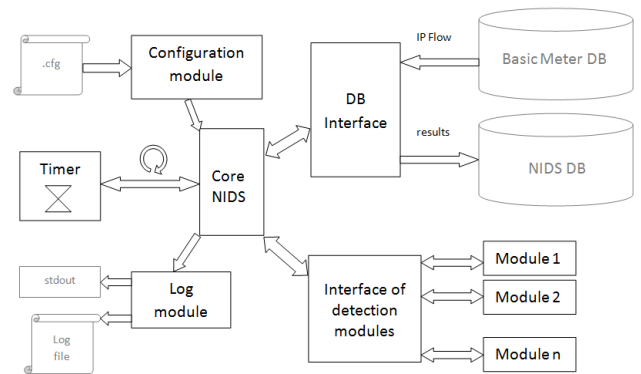


Fig. 2: Modules of the detection system

modules where each module consists of a group of its own source files. Core structure is shown in Fig. 2.

The detection system was developed as a modular. Modularity is a significant opportunity for potential future expansion of the proposed system. Emphasis was also placed on creating a simple interface for modules, which allows very easy way to create new extensions.

- Configuration module - access configuration data from file.
- Database interface - represents interface to database access of BasicMeter and database of IDS system.
- Time planner - time planning of network traffic monitoring
- System core - interconnect every module of system. Main task is initializing modules cooperation.
- Detection module - represents layer between system core and detection modules. Designed for simple adding and removing modules to/from the system.

A. SYN Flood Attack Detection

One of the most common Denial of Service attack is called a SYN flood. SYN flood attack is designed to overload services that use TCP communication channel [4]. In this type of attack is used substance of establishing communication in the TCP connection.

Communication in SYN flood attack (Fig. 3):

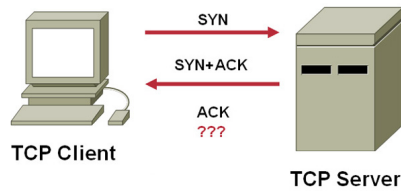


Fig. 3: Communication in SYN flood attack

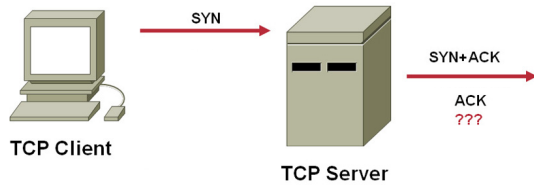


Fig. 4: Communication in SYN flood attack with non-existent IP address

- 1) Client sends a TCP packet with SYN flag to server. Packet may contain non-existent IP address of the applicant links (Fig. 4).
- 2) Server sends a packet with SYN and ACK flags to IP address contained in the packet
- 3) Server awaits ACK packet, but that does not receive
- 4) After expiry of the time allowed for binding of communication (timeout) server releases memory resources. Timeout can take tens of seconds to minutes, according to server settings.

If an attacker manages to send a large number of fake packets, to time runs out of memory reserved for the operating system TCP connections and causing rejection of requests for connection, it is a requests from regular users.

Some SYN flood attacks handling the IP address, but after the adoption of SYN-ACK packets are not transmitted packets with the ACK flag.

SYN flood attacks are based to sending a TCP packet with SYN flag and usually with forged source address. IPFIX platform counts TCP flags that occurred in the flow. In the traditional TCP connection is tied to the number of received SYN flags less than or equal to the number of ACK flags. The attack SYN flood are generated flows, which contain more SYN flags such as ACK flags, usually one symptom of SYN and zero ACK flags.

Detection module detects that the flows contain more SYN flags such as ACK flags and that difference count in the specified interval. If this value exceeds the limit, the module is considering this state as a SYN flood attack. For this attack detection is not possible to identify the attacker based on data from IPFIX protocol.

B. Detection of UDP Flood Attack

UDP protocol [5] is a communication protocol without connections. UDP protocol does not guarantee that the datagram is delivered, served several times, or datagrams are transmitted in a different order than they were sent. When the UDP flood attack forges the source address, so answer is sent to another computer as the source and packet filtering based on source addresses is ineffective.

UDP Flood DoS attack forms UDP packets of different sizes, different content and with falsified source address.

Detection principle is based on counting the number of bytes of incoming UDP packets transmitted. If the sum of data in the specified range exceeds the limit, can be regarded as a potential attack. Similarly as in the detection of SYN flood attacks, in this attack is not possible to obtain data from IPFIX information about the attacker too.

UDP flood is one of the least dangerous, as there are mechanisms to effectively defend him. Reliable defense are firewalls in key areas of network, which filters out unwanted communications. For larger attacks, the solution is to ask the Internet Service Provider for packet filtering.

C. Detection of HTTP Flood Attack

HTTP flood is focused to standard web servers. The essence is in sending a large number of HTTP requests [6]. HTTP flood usually focuses to the web site, which disproportionately used sources like database server, which can shut down a Web service lines because of congestion of HTTP server or database server. HTTP flood attack is one of the most widely used assault for its simplicity and destructiveness.

HTTP flood attack sends the number of HTTP requests that are attacking to limits of some of the HTTP server layers. In this type of attack is not used forged source IP address and offensive content HTTP requests has the same structure as a legitimate request. For information IPFIX protocol can not be distinguished on what the request is. Detection of HTTP flood attacks based to calculating the incoming flows of TCP transport protocol with the target HTTP server port (usually port 80). If the sum of flows in the specified timeout exceeds the limit, the detection module notify the HTTP flood attack.

If the attack with a single attacker can obtain its public IP address such as finding addresses, from which came the largest number of requests. In a distributed (DDoS) attacks to trace the attacker is not possible.

There is no effective defense against HTTP flood. For smaller attacks defense lies in the blacklists of IP addresses (IP blacklists) and increasing capacity web server. For larger attacks is possible to establish a defense to analyze packets in which the search for common features (masks) HTTP requests.

D. Detection of Some Other DoS Attacks by Measuring The Latency

The last method is based to completely different principles than the previous one, which tried to detect a specific type of attack it directly from symptoms of the network operation. The data obtained by IPFIX system can calculate the delay response to a request asking for a connection. This value may reflect the effects of several types of DoS attacks, as well as local DoS attacks. For the calculation of latency is needed to get two streams belonging to one (bidirectional) TCP connection. Latency is the difference times of the beginnings these flows.

The disadvantage of detecting DoS attacks this method is the possibility of "False Positive" events during maintenance of the system during normal operation.

E. System database

Database of implemented intrusion detection system is based on MySQL DB. Relation model consists from five tables Fig. 5. Main tables is Intrusion, which stores information about detected intrusions. Addition tables stores historical data of monitored computer network.

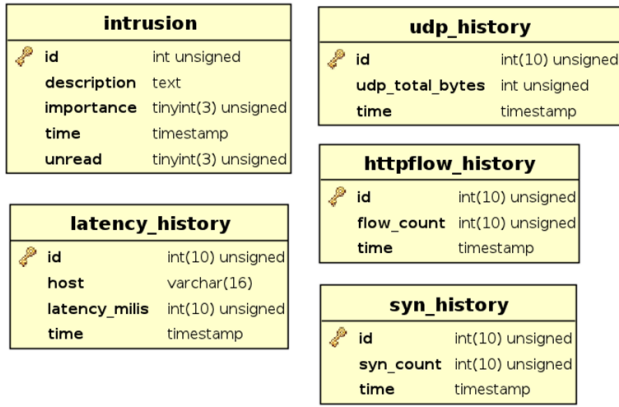


Fig. 5: Relation model of detection system

TABLE I: System configuration

Configuration	Server	Attacker
Procesor	Intel Core2duo, 1.8GHz	AMD Sempron 2800+
RAM	4GB	768MB
HDD	160GB	80GB
OS	Ubuntu Linux 11.04 Beta	Ubuntu Linux 10.10
Kernel	2.6.38	2.6.38

III. VERIFICATION OF DETECTION SYSTEM

The detection system was tested on a computer network with two computers: First, an intruder's computer, and second, a server with the installed detection system and running regular services, such as httpd, ftpd, or sshd. The system configuration is stated in Tab. I. The computers were connected with a network Fast Ethernet 100BASE-TX.

All the testings were performed in two steps: First step took 15 minutes monitoring of the standard traffic parameters. Second step involved enabling of the particular attack type in order to track the network's parameters during the attack. For each attack type, the tracked values were displayed a short time before the attack launch and a while after it had started. The graphs reflecting quantity of the UDP bytes, http connections, and suspicious SYN packets, contain values measured within 10 seconds intervals.

A. Detection of HTTP Flood Attack

Detection of HTTP flood attacks is based on counting of http connections. As the attacking application, a simple program has been used. It is programmed in the bash scripting language, and within the loop it is sending http requests:

```
#!/bin/bash
while true; do
    wget http://10.0.0.2/index.htm
    rm index.htm
    sleep 0.1
    echo #
done
```

In Fig. 6, number of the http connections is shown in the 15 minutes interval during normal operation. In Fig. 7, high growth of the http connections is reflected after the launch of the http flood attack.

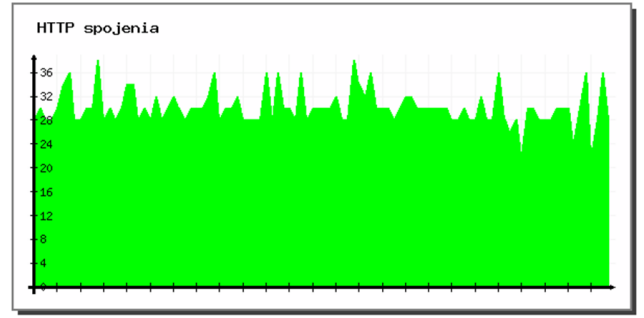


Fig. 6: Time course of http connections during normal operation.

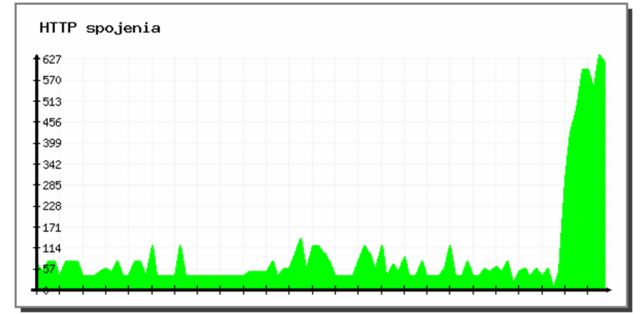


Fig. 7: Time course of http connections during http flood attack.

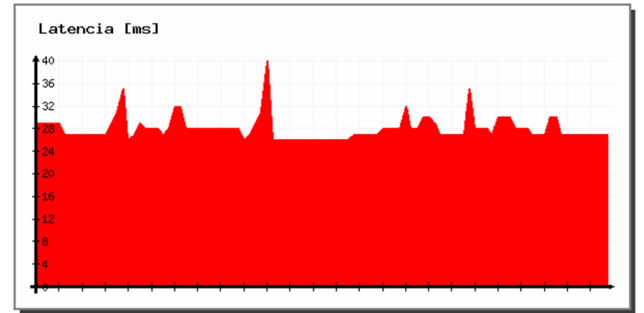


Fig. 8: Time course of response to connection during normal operation.

B. Detection of DoS Attacks via Latency Measure

The last tested method was latency measure of the response to a http connection request. During normal operation Fig. 8, the maximum response approximated at about 30 milliseconds. As the attacking program, a local DoS attack was performed, fork() bomb (). Shortly after the start of the attack, the latency acquired values around 600 milliseconds. Subsequently, operating system stopped responding.

IV. CONCLUSION

The aim of this work was the design and implementation of the detection system of DoS attacks. Network traffic were obtained from the measurement platform trunk IPFIX protocol. This platform provides the necessary amount of data to detect the most common DoS attacks. Detection system was designed as modular, with the possibility for a simple creating of new detection modules. Modules were implemented to detect the three types of attacks: HTTP flood, SYN flood and UDP flood. In addition to these modules was created a fourth, which seeks to detect several types of attacks, measuring the time of the

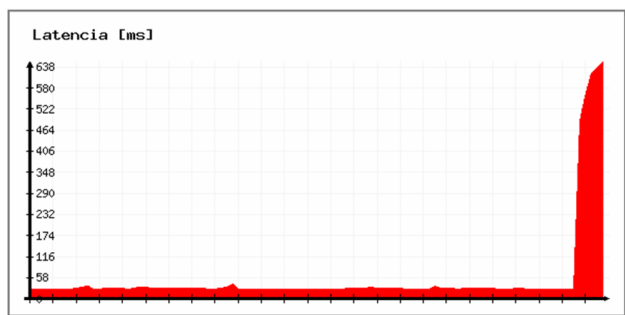


Fig. 9: Time course of response to connection during local DoS attack.

system response. Testing demonstrated the ability to detect some types of attacks. During implementation and testing were used only open and freely available software packages including operating system, libraries and the DoS attacks.

ACKNOWLEDGMENT

This paper was developed with support of the Project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky" (Centre of Excellence of Integrated Research and Exploitation the Advanced materials and Technologies in the Automotive Electronics), ITMS 26220120055, that is co-financed from Structural Funds EU ERDF within Operational Program Research and Development OPVaV-2009/2.1/03-SORO and preferred axis 2 Support of Research and Development.



REFERENCES

- [1] The Internet Engineering Task Force, Fremont, California: RFC 791: Internet Protocol. URL: <http://www.ietf.org/rfc/rfc791.txt>
- [2] Network Working Group: Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information. URL: <http://tools.ietf.org/html/rfc5101>
- [3] CNL KPI FEI TU Košice: MONICA Research Group. URL: <http://www.cnl.sk/projekty/monica>
- [4] The Internet Engineering Task Force, Fremont, California: RFC 793: Transmission Control Protocol. Dostupn na internete: <http://www.ietf.org/rfc/rfc793.txt>
- [5] The Internet Engineering Task Force, Fremont, California: RFC 768: User Datagram Protocol. URL: <http://www.ietf.org/rfc/rfc0768.txt>
- [6] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach a T. Berners-Lee: RFC2616: Hypertext Transfer Protocol - HTTP/1.1. URL: <http://www.ietf.org/rfc/rfc2616.txt>
- [7] Ádám N. - Single input operators of the DF KPI system. In: Acta Polytechnica Hungarica. Vol. 7, no. 1 (2010), p. 73-86. - ISSN 1785-8860
- [8] Tomášek M., Čajkovský M., Madoš B. Intrusion detection system based on system behavior 2012. In: SAMI 2012 : 10th IEEE Jubilee International Symposium on Applied Machine Intelligence and Informatics : proceedings : Herany, Slovakia, January 26-28, 2012. - Budapest : IEEE, 2012 S. 271-275. - ISBN 978-1-4577-0195-5

Design and Development of Application for Supervisory Control

¹Lukáš LACIŇÁK (1st year), ²Roman MIHAL (2nd year)
 Supervisor: ³Iveta ZOLOTOVÁ

¹Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic
^{2,3}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹lukas.lacinak@tuke.sk, ²roman.mihal@tuke.sk, ³iveta.zolotova@tuke.sk

Abstract— This paper deals with the creation of applications for the supervisory control model Flexible Production System. The work aims to describe the creation of applications by using the Intouch software tools and application server. It emphasizes the use of object-oriented approach and modern graphic symbols in a development environment ArchestraA, as well as describes network architecture on which the application will be created and deployed.

Keywords - supervisory control, SCADA/HMI, visualization, screens.

I. INTRODUCTION

Applications for supervisory control are used in all sectors of industries. Their main objective is to monitor, track, and control real processes in plants. When creating applications designed for supervisory control it is necessary to consider many factors [1].

The first part of this paper focuses on analysis. Draft layout screens are created in accordance with customer requirements, which consist of individual windows, design of windows and their hierarchy, and last but not least, the essential objects that represent the real process.

The application for supervisory control of model FPS (Flexible Production System) at the system level SCADA/HMI (Supervisory Control and Data Acquisition / Human Machine Interface) [5] is described in the second part of the document. This brief description deals with application and its functionality. Applications of such significance should meet several criteria, such as transparency, intuitiveness and ease of readability for the operators working with this application.

The final section describes network architecture on which this application was created and then also deployed. This architecture contains all the resources that were used to create applications for supervisory control of the FPS model. This architecture corresponds with the architectures used in real production [3].

Application is created as the output of the thesis. The aim of this study was to develop an application for control of real supervisory FMS model. This paper deals with the description of the application itself.

II. DESCRIPTION OF MODEL FPS

The model of FPS represents a laboratory model of the production line intended for sorting cubes, placing pallets according to the order and placement of orders on the shelf. It consists of 6 posts [3].

At the first post, the implementation of templates based on customer orders from a software application called IS (information system) takes place. Individual blocks are being placed on a pallet by a triaxial manipulator with a suction disk. Cube lifting and laying is done by a pneumatic piston. After the cubes are placed in pallets in the pattern, the pallet moves to post 2 [8].

Post 2 carries out a camera check-up and passes the template between conveyors. A bar code is located at the bottom of the pallet and after scanning this code, the IS can provide information about the pallet in this position concerning the content, i.e. the pattern to be placed. By comparing the scanned pattern placed on a pallet with the information from the IS about the requested pattern, the concrete pallet is being evaluated either as placed correctly or incorrectly (scrap) [3], [7].

If the pallet was placed correctly, it continues to post 3. At this post the pallet is stowed into the rack, while the movement is carried out by the three-axial manipulator with jaws for gripping the pallet. Information about the location of placed pallets is sent into the IS. When the order is placed, all pallets for the given order are removed from the rack [3].

These are then gradually transported by the second conveyor to post 4, where the pallet is moved back to the first conveyor. After this, the pallet is moved to the post 5. Also pallet evaluated as a scrap on post 2 is transferred from camera control to the post 4, where transfer between conveyors is being carried out [3]. After this conveyor, the pallet is moved to post 5, where emptying pallets takes place. The pallet is gripped by jaws on the shoulder, which then empties the contents of the pallet into a vibrating container.

The first step is sorting the cubes of four colors (white, green, blue, black), which are moved from the vibrating cartridge to the sorter by a small conveyor. A scanner is placed above this conveyor, which detects the color of cubes and those are then sorted into small cartridges. In the event a

cube can't be recognized, it is sent back into the vibrating cartridge on a small conveyor [8].

Empty pallet is sent back to the start, i.e. post 1, where the filling of templates is carried out. The oversight of the model and arrangement of posts is depicted in Fig. 1.

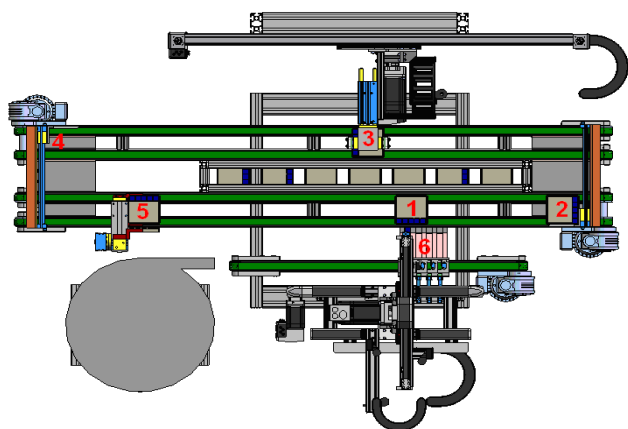


Fig. 1. View model of the Flexible Production System

III. ANALYSIS AND DESIGN OF THE APPLICATION

When creating an application, a great number of details need to be taken into consideration. The first factor is screen layout of the application [3]. This screen consists of several windows, which are displayed in different parts of the manufacturing process. Distribution of the main screen is shown in Fig. 1. The application consists of three windows:

- Upper menu
- Left menu
- Control Window

The upper menu allows us to get to the window, which displays alarms and trends. It is possible to get to the login window, if the currently logged user wanted to log off or change the password, or log in as another user [7].

The left menu allows us to switch between windows for specific positions and the preview window.

The control window is used to visualize different parts of the process and to display window alarms and trends. This screen is divided into three parts, namely:

- Name and status
- Control elements
- Elements of visualization

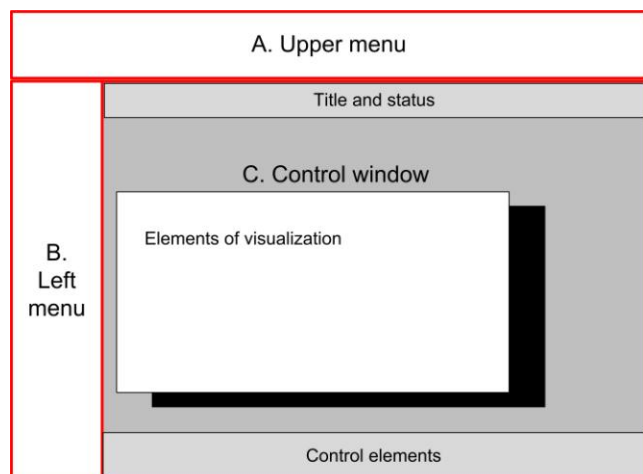


Fig. 2. Distribution of the screen [3]

Finally, it is necessary to mention the unification of colors. These colors have been standardized for the active components such as conveyors, vibrating container etc. These elements have a unified color, which means that if the active element is shown in green and if it is inactive, i.e. when the conveyor is turned off, then it is displayed with blue color.

The next step was creating a hierarchy of windows. This hierarchy gives us information about which windows are displayed in what order and how it is possible to switch between windows, Fig. 3.

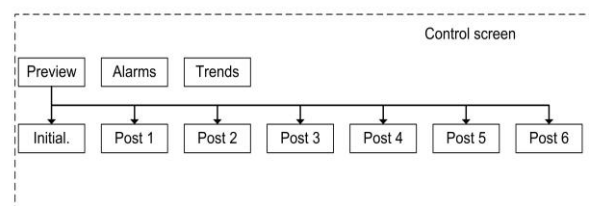
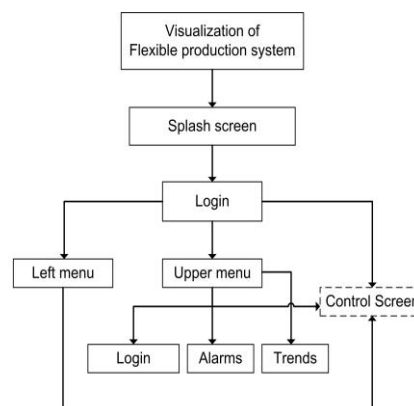


Fig. 3. Architecture windows for visualization [3]

IV. DESCRIPTION OF THE APPLICATION

After launching the application the splash screen is displayed, followed by the login window. In the login window, the login information for the user who wants the current login can be specified. This provides an option to log in, log out, or change the current user's password. If an unrecognized user is trying to log in, he will be logged in as a visitor [3].

After login, the screen layout as in Fig. 3 can be seen. This screen consists of the three windows mentioned before.

As the screen is divided into several parts, each of these parts fulfills a specific function. Parts A and B, according to Fig. 2, are always visible after logging in. Part C, the control box, is designed to display various screens. Figure 4 shows the screen that appears immediately after logging in.

Part A (upper menu) allows us to switch to windows with alarms and trends and to show the login window. It displays the currently logged user and notifies him in case of long inactivity. The current time serves as an information element. This menu also displays a warning if the alarm situation occurs, which signals the logged user to check alarms, to determine what error occurred and to resolve the situation [3].

Section B (left menu) is used to switch between individual previews of the application and the general overview.

The most important part is C, where it is possible to observe the ongoing activities of the individual positions. This section is further divided into other areas, so that individual windows

displayed in this section are normalized in the same way.

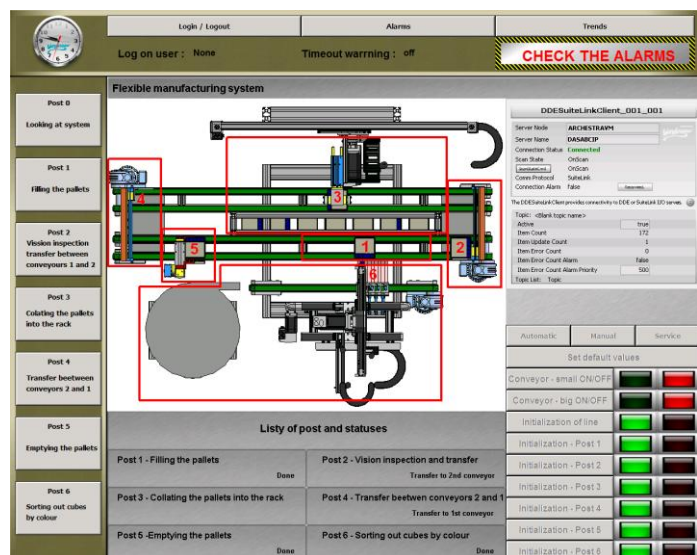


Fig. 4. Splash screen

Figure 4 clearly describes that the control window is divided into the following sections.

1. Contains information about title of the post and displays the status of this post, i.e. information about what activities are currently ongoing.
2. This represents a control section, which includes individual control elements. After logging in as supervisor, these elements can be used to control the individual elements of the line.
3. Contains actual visualization. Objects created in the graphical environment ArchestraA. These objects are graphically illustrating the various activities going on the posts.

All windows were normalized this way. The only exception is the preview window, in which controls are placed in a different way [6].

Screen for post 3 can be used as an example. This screen includes setting up the pallet into the rack with all the necessary information and control elements, as shown in Fig. 5. The window consists of several control and signaling elements.

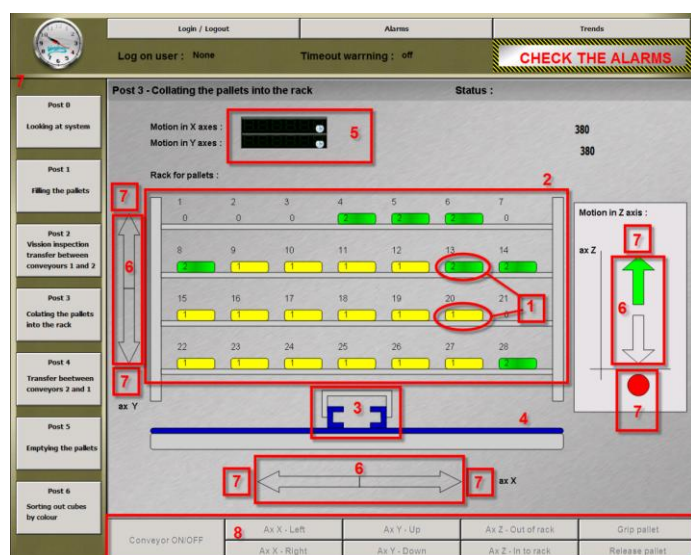


Fig. 5. Screen for post 3

Legend and description of individual elements:

1. **Pallets** – there is 28 of these elements, each represents a variety of presence or absence of a particular place in a rack. It also displays the bar code of the pallet.
2. **Rack** - a graphical display rack. It is a static element without defined features.
3. **Jaws** – represent an active element. They include a pallet to start out in the rack.
4. **Belt** - this is a conveyor to move pallets between posts. Indicates to us that the conveyor is turned on or off.
5. **Displays X, Y** - the two displays continuously provide information about the current position of the manipulator arm axes X and Y.
6. **Signaling movement** – arrows were used as an element for display.
7. **Limit sensors** - these elements are normally hidden. They are activated after the arrival of the arm in extreme positions. The Z axis is always one of the two active sensors, since this axis arm can only take two extreme positions.
8. **Control elements** - these are active when the user logged in administrator and switch into service mode. They provide control of individual elements of the model.

V. NETWORK ARCHITECTURE

The application was created and deployed on the network architecture shown in Fig 5. This architecture is based on the principle of client-server architecture. Each server provides the services of a specific character [2]. The advantage of this architecture is that different services are independent and in case of outage, only one service will be affected. Another advantage is ease of configuration and maintenance of the entire application [5].

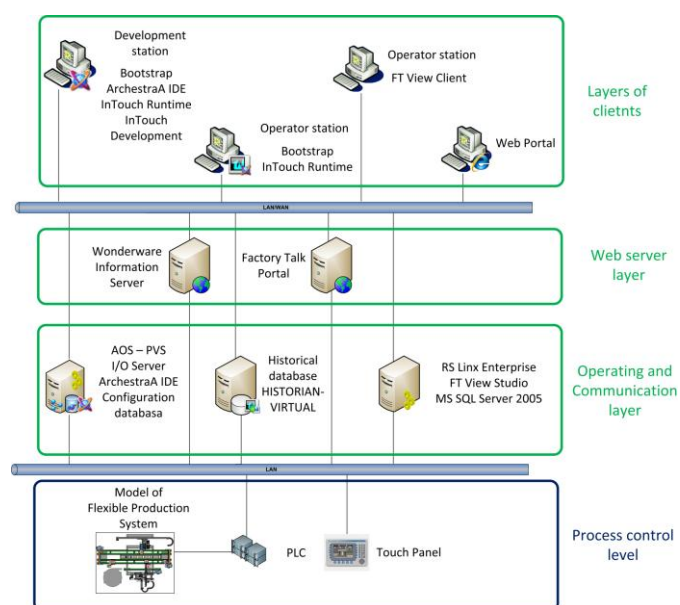


Fig. 6 Client-server network architecture

VI. CONCLUSION

The purpose of this document is to show and describe the creation of applications for the supervisory control model FPS. The described procedure is used to design and develop

new applications in the laboratories of the Department of Cybernetics and Artificial Intelligence. The described application is used to control the model in laboratory conditions. Along with the development application, the network architecture on which the application was deployed is developed and reshaped.

The next step in extending the application would be adding new controlling and monitoring elements (such as monitoring the model by camera in real time) and many other tasks related to the model.

ACKNOWLEDGMENT

The work presented in this paper was supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic within the KEGA 021TUKE-4/2012 project (70%) and VEGA - 1/0286/11 project (30%). This work was concluded in our laboratory with the software and licenses provided by Wonderware Educational Consignment Support from Pantek (CS) s.r.o.

REFERENCES

- [1] A. Daneels, W. Salter, "What is SCADA?", International Conference on Accelerator and Large Experimental Physics Control Systems, Trieste, Italy, 1999, pp. 339-343.
- [2] J. Flochová, L. Mraňko, D. Mudrončík, F. Auxt, "On Industrial Techniques and Information Technology Control Classrooms", Proceeding of IEEE Conference on Decision and Control'07. New Orleans, 2007, pp. 1874-1879.
- [3] L. Laciňák, Supervisory Control Level of Flexible Production system, Diploma thesis, Košice, 2011, pp. 85.
- [4] L. Landryová, M. Babiuch, "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.
- [5] I. Zolotová, M. Pavlík, R. Mihaľ, M. Kollár, "Vzdialené monitorovanie a supervízne riadenie na báze aplikačného a webového server", Electrical Engineering and Informatics 2, pp. 532-535, Košice FEI TU, 2011, pp. 532-535.
- [6] I. Zolotová, M. Bakoš, L. Landryová, "Possibilities of Communication in Information and Control Systems", Annals of the University of Craiova, vol. 4, no. 2 (2007), p. 163-168
- [7] I. Zolotová, L. Laciňák, S. Laciňák, "Machine and site supervisory control of flexible production system", International conference on applied electrical engineering and informatics 2011, Košice : TU, 2011 pp. 27-31.
- [8] I. Zolotová, L. Laciňák, R. Mihaľ, "Data acquisition and supervisory control – communication layer and life cycle of application", CD-ROM, Bratislava : STU, 2012 .
- [9] I. Zolotová, M. Pavlík, "Building a High Availability Solution in Laboratory Environment", Management Science and Engineering, Chengdu, China, 2011, pp. 621-625.
- [10] I. Zolotová, M. Pavlík, R. Hošák, L. Landryová " Benefits of virtualization in HMISCADA systems", System Theory and Control, Sinaia, Romania, 2010, pp. 379-384.

Edukit: A Real-Time Web-Based Collaboration Tool for Educational Purposes

Vojtech RINÍK (3rd year Bc.), Ivan KLIMEK (2nd year PhD.)

Supervisor: Assoc. Prof. František JAKAB, PhD.

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

vojtech.rinik@student.tuke.sk, ivan.klimek@tuke.sk, frantisek.jakab@tuke.sk

Abstract—This paper describes Edukit: A web-based tool that allows information sharing by providing features for sending and receiving messages. As the main strong point Edukit offers a feature that, in effect, transforms communication data into structured pages. The system is aware of messages with important knowledge, it extracts data from the communication, archives it, classifies into categories, and facilitates access to it. Additionally, it implements very fast, real-time user interface based on local storage and data synchronization.

Keywords—collaboration tools, knowledge, information sharing

I. INTRODUCTION

Online user collaboration is a very common problem that various software products try to solve. There is over a hundred of tools listed in the related Wikipedia page. [1]

Most of these tools serve only as a medium of getting information from one user to another. They are not aware of what do the data mean, and if they do, it's only for purpose such as filtering spam emails. [7] The main idea of Edukit is to inspect the messages that are sent and received using the tool and find knowledge in them. Next step is data classification, which is finding data that are related, and putting them into same groups. Finally, the last step is to provide an easy access to extracted knowledge.

Another problem with some of the existing solutions is the lack of real-time experience. With the modern browsers, there is a possibility to add real-time features to every application.

The third issue Edukit attempts to address is the current state of web user interfaces. Most of the web interfaces seem slow and awkward because user is always notified when there's a network transfer going on. Edukit utilizes Atmos2 [2], a framework that caches data locally and synchronizes them with the server in background.

Edukit is an application for group collaboration, with intelligent knowledge extraction and classification and a fast real-time web interface. The following sections will discuss each one of these aspects.

II. "INTEL" MINING

Edukit provides an interface for sending and receiving messages, similar to group email. This section discusses "Intel mining", a process that selectively finds and stores useful information contained in the communication data.

The input is a collection of messages; each one is processed by an evaluation procedure that decides how useful it is. The basic principle of this processing is depicted in Figure II-1.

Message stream	
Hello everyone	score: -0.2
The password is ...	score: 6.4
Here's the link with ...	score: 7.2
Thanks!	score: -1.4

Figure II-1: Message stream and computing score

The result of message processing is a score, which is computed based on several input variables. The first one is number of votes from other users. Voting up single pieces of information is a concept that users are familiar with thanks to popular Facebook feature, "liking" posts and comments. [5] It is also a form of collective intelligence. [9] Another variable is the age of message. It's evaluated in a way that messages that were posted recently and have received many votes are the most useful ones. The last one is the content: instead of lexical parsing, the importance is counted from metadata such as length, presence of hyperlinks or presence of code blocks.

The computation of individual score is illustrated in Figure II-2: Factors that affect score of a message.



Figure II-2: Factors that affect score of a message

After this processing there are scores for each messages, the ones with score high enough are converted into so-called cards.

Cards are small pieces of information. A card can be created from a message automatically, or a user can add it manually. User can also edit a card correcting possible mistakes in the original message.

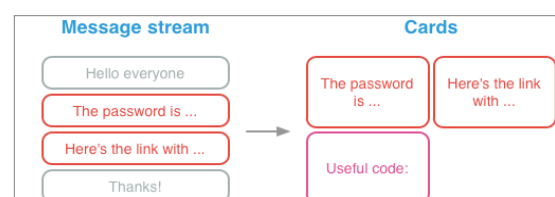


Figure II-3: Cards created from messages or created manually

III. INTEL CLASSIFICATION

In real world scenario, cards may concern many different topics. To avoid confusion of displaying all cards at once, cards can be added to stacks as illustrated in Figure III-1: Creating stacks from cards

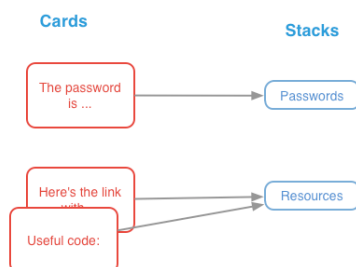


Figure III-1: Creating stacks from cards

The first way of classifying cards is moving them manually to the correct category. After some of the cards are classified, the system utilizes learning algorithm to automatically decide where to put new cards.

The way this works is by computing similarity score between two cards. The algorithm is similar to the nearest neighbor algorithm. [10] When a new card needs to be classified, existing, classified cards with high similarity score (with short distance in terms of nearest neighbor algorithm) are looked up, and if they are found, the card is added to the relevant folder. User can always manually move card to a new stack.



Figure III-2: Similarity score between two cards

The method of counting similarity score is illustrated in the Figure III-2: Similarity score between two cards. This algorithm is rather simple and it takes things like what thread were the original messages part of, or how soon one after another were posted. Also, the contents are parsed and compared for similarities between the two messages. In this processing, the structure of the document, written using Markdown formatting [6] is taken into account: For example, text in the titles is considered more important than text in the paragraphs, so if there's the same word in two cards in the titles, the similarity score is higher than if it were in the paragraph.

IV. CACHED, REAL TIME USER INTERFACE

Edukit uses two innovations in the area of user interfaces. First is caching of all objects: Traditional “downstream caching” - when user opens the application, objects from local cache are loaded while current objects are transmitted over the network. [4]

Novelty of the presented approach lies in its “upstream” caching meaning when user creates a new object, it's stored in the local cache and synced in the background with the server without blocking further user activities. This results in what seems to be an instant user interface, user is never blocked, never has to wait for load to finish and the overall feeling resembles to using a desktop application.

The second innovation is real-time user interface. Its first part is displaying presence of other users. For example, when viewing list of projects, the user can see little avatars of their

colleagues next to the project list. This resembles to instant messaging protocols, but is included in the web interface of Edukit.

Edukit also includes collaborative editing of cards. To avoid conflicts, operational transformation [3] [8] is applied when multiple users edit the same message. The result of applying OT is that there is never a situation when two users might see different content for one document.

The principles of OT implementation are described in Figure IV-1: Operational Transformation Considering that two clients are editing the same document, and they both type a character at the same time, there is a situation that needs to be resolved to avoid conflicts. Operation transformation transforms the two operations in one operation that produces the same result in both clients.

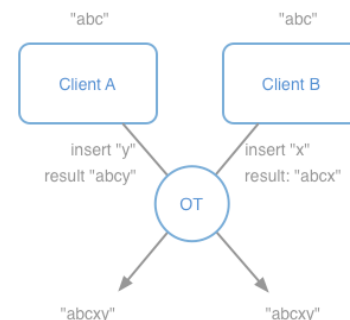


Figure IV-1: Operational Transformation

V. BETA TESTING DEPLOYMENT

Edukit is not only a concept, it's a platform that's implemented, deployed and currently used by about 200 students at Technical University of Košice. During the winter semester of academic year 2011/2012, these users have generated 12362 visits and 62132 pageviews in total. A total of 1003 messages were posted. It is used mostly by a closed group of students because information in Edukit is relevant only to them. The problem of expanding to other schools is still open.

The following results are concluded from verbal feedback and analytics of pageviews.

One of the strong points discovered during the beta is the interface for accessing information in form of pages. Once the page is created and filled with information, user prefer reading a well-structured page rather than finding information in old messages.

The second strong point is the messaging module. It provides an interface with all of the messages from a course and the unread messages are highlighted. Posting a new message is almost instant thanks AJAX.

The results also imply that there is a small number of users who actively contribute by editing pages and creating new ones. Most of the users prefer to communicate – write and read messages. This trend is illustrated in Figure V-1: Usage statistics of Edukit modules.

Simply put, most of the users like to read well organized documents with information, but don't like to create them. Still they contribute with information by sending messages.

This trend has demonstrated that there is strong need for implementing the functionality described in this paper: If the system could find and classify the knowledge in a stream of messages, a lot more of knowledge would be accumulated, overall improving the usefulness of the tool.

Additional fact learned from the beta run is that people come back to consume information in the system especially before the time of exams. That shows it's very important to have an easy way to access knowledge in an organized manner even some time after it was entered into the system.

Figure V-1: Usage statistics of Edukit modules illustrates the actual statistics of accessing the messaging module of the system, versus the pages module.

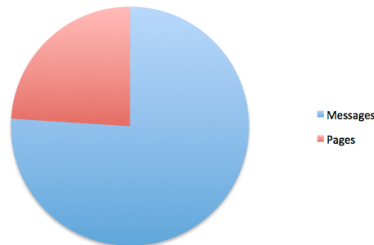


Figure V-1: Usage statistics of Edukit modules

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] WIKIPEDIA: List of collaborative software [Online] Available: http://en.wikipedia.org/wiki/List_of_collaborative_software
- [2] V. RINIK: Atmos2, a JavaScript synchronization library [Online] <https://github.com/vojto/atmos2>
- [3] D. SPIEWAK: Understanding and Applying Operational Transformation [Online] <http://www.codecommit.com/blog/java/understanding-and-applying-operational-transformation>
- [4] B. WATTERS: Developing RESTful iOS Apps with RestKit [Online] Available: http://mobile.tutsplus.com/tutorials/iphone/restkit_ios-sdk/
- [5] FACEBOOK: Like [Online] Available: <http://www.facebook.com/help/like>
- [6] J. GRUBER: Markdown [Online] Available: <http://daringfireball.net/projects/markdown/>
- [7] D. MERTZ: Spam filtering techniques [Online] Available: <http://www.ibm.com/developerworks/linux/library/l-spamf/index.html>
- [8] D. WANG, A. MAH, S. LASSEN: Google Wave Operational Transformation, Google, July 2010.
- [9] T. AITAMURTO, A. LEIPONEN, R. TEE: The Promise of Idea Crowdsourcing – Benefits, Contexts, Limitations, Nokia, June 2011.
- [10] P. HARRINGTON: Machine Learning in Action, Manning, April 2012.

ETL System Proposal

¹Eva DANKOVÁ (3rd year), ²Adrián PEKÁR (1st year)
Supervisor: ³Liberios VOKOROKOS

^{1, 2, 3}Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹eva.dankova@tuke.sk, ²adrian.pekar@tuke.sk, ³liberios.vokorokos@tuke.sk

Abstract—This paper handle about data warehouses, ETL systems and its proposal. In first part of this paper is a theoretical description of data warehouse and its architecture, which is a basis for the proposal of the ETL system. The second part handle about proposed architecture of the ETL system, which should be used for recording data from heterogeneous databases and their offline extracts into data warehouse.

Keywords—data warehouse, data mart, extracting, transmitting, database.

I. INTRODUCTION

Mass use of information technology in all sectors of human life, is reflected on the amount of data that accumulate in Information Systems. The accumulation of data implies the question: „how can be these data processed, stored and used primarily for the benefit of their owner“. Here comes to the scene database systems that can store huge amounts of information, which can be available for users through information systems. However, if this data have to be worked out, they need to be processed into a form in which they would be easier and more efficient to work with. This form serve us data warehouse, which is built to store amount of information and to make data more effective by analysis. This analysis helps to decide by making business.

To get data from multiple sources, which may differ in database structure or in used technology, to data warehouse it is needed a tool, which allows uniting the data. For these are used ETL systems, that means extraction, transformation and store of data in warehouse. There are many manufactures of these systems and many of them provide us a complex solution suitable for most of data warehouses. This paper handles about methodology of designing ETL system and the design of ETL system.

II. DATA WAREHOUSE

Data warehouse is an enterprise structured depository of subject-oriented, integrated, time-varying and historical data used to obtain information and decision support. In data warehouse are stored atomic and summary data. [4]

Data are acquired from production (operational) databases, which can be in different departments of companies or in different geographic locations. These data are in regular intervals collected prepared for process and introduced into the data warehouse. A data warehouse is essentially also a database, which is organized by other type of rules. From this rules imply also other properties (tables may not be

standardized). Data warehouse is a set of technologies for efficient data storage, so that data converted into information are used for making decision. [4]

TABLE I.
COMPARING DATA WAREHOUSE AND OPERATIONAL DATABASES

Properties	Operational database	Data warehouse
Response time	Fraction of seconds	Seconds to hours
Operations	Data manipulation language	Primary only to read
Data origin	30-60 days	Snapshots sequence in time section
Data organization	Based on application	Based on subject, time...
Size	Small to big	Big to huge
Data resources	Operational, internal	Operational, internal, external
Activities	Processes	Analysis

A. Basic architecture

The basic architecture of data warehouse is shown in Fig. 1. This architecture is theoretical and in praxis is a basis for other architectures. Data source for warehouses are production systems or other resources (file in spreadsheets, XML files, and external data tables, etc.). On the other site are users and systems that receive data from the data warehouse. Parts of data warehouse are row data (classical tables' dimensions, facts, etc.), metadata (object description) and aggregate objects (tables containing data aggregation). [9]

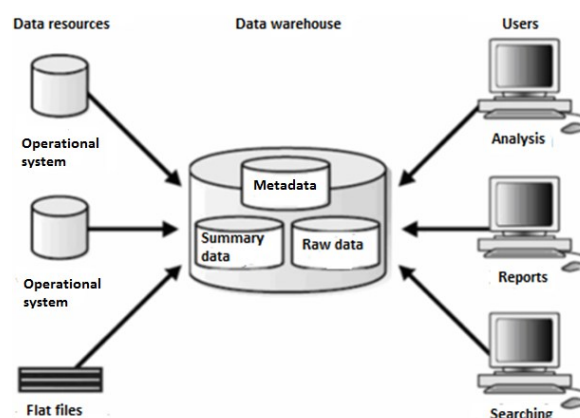


Fig. 1 Basic architecture of data warehouse

B. Data mart

The latest improvement in data warehouse architecture is the addition of so-called data mart places (Fig. 2). According to definition of “data mart” are precisely specified subsets of the data warehouse, which are for smaller branches of the companies. [9]

Data for data marts are collected to meet the specific needs of organization parts. Data mart is often preferred by companies as a first step to build a data warehouse. It can be also used as proof of the data warehouse concept correctness. Data mart is a special version of the data warehouse, which also contains images of operational data.

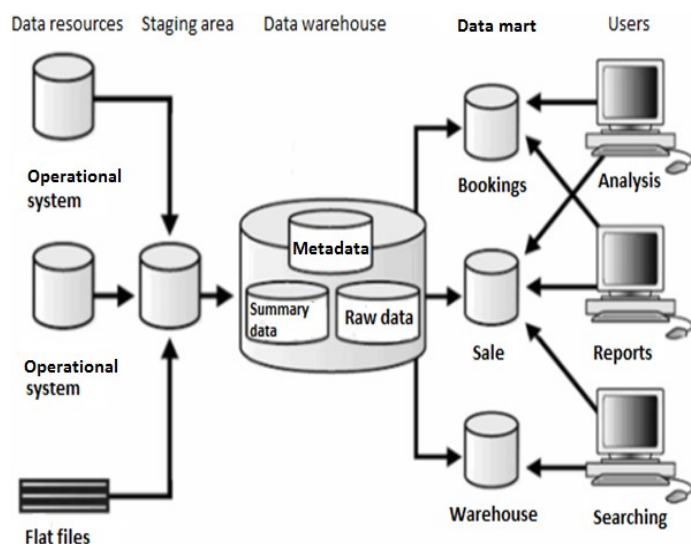


Fig. 2 Architecture with use of staging area and data marts

The main difference is that by creating the data marts focuses on specific, predefined needs of users and data configuration. Data stored in marts can mainly be used as a basis for targeted analysis.

The advantage of this arrangement in more separate data warehouses is easier and faster implementation. So the resulting benefit to the user is faster access. On the other hand, the disadvantage is that there may be inconsistency between different marts. This scheme is also difficult to maintain. [9]

III. METHODS OF BUILDING WAREHOUSE

Probably the most important step in building a data warehouse is to select the most appropriate method. We must take into account not only the structure organization and information “culture” of the company, but also forecast of possible problems, which can occur during the construction of data warehouse. The best known and most common methods are:

- “Big bang” method
- Incremental method

A. “Big bang” model

Development of a data warehouse is a complex issue and it probably will not be settled at once and certainly not within a reasonable time.

This is the greatest weakness, because if a data warehouse project was finally by the big bang method implemented,

meanwhile may change technology and user requirements. Big bang method consists of three stages:

- Requirements analysis
- Creating of data warehouse
- Creating access to data warehouse

B. Incremental model

Incremental method, otherwise known as evolutionary, expects building of data warehouse in several stages. It means that instead of building the whole store, it gradually accrues solutions that fit naturally into the overall data warehouse architecture. [10]

This partial solution can be implemented such as a scalable data mart and is provided to end users and thus partially satisfy hunger for return of inventions. So first subsystems start to work and bring benefits in a short time.

If a partial solution tested by users certified, it can be added next area, so it gain a new functionality. And so it could continue until a finally resolution of the data warehouse will be build. Building a data warehouse is therefore an interactive process, which still maintains a link between the data warehouse and user needs. The main advantages of the incremental method include [10]:

- Incremental development of data warehouse maintains continuity of the project with the needs of users
- It allows to implement a scalable, extensible architecture
- Provide faster gain and return on investments.

If was decided to build a data warehouse using the increment method, we can choose between two options:

- Incremental method toward “top down” (Fig. 3)
- Incremental method towards “bottom up”

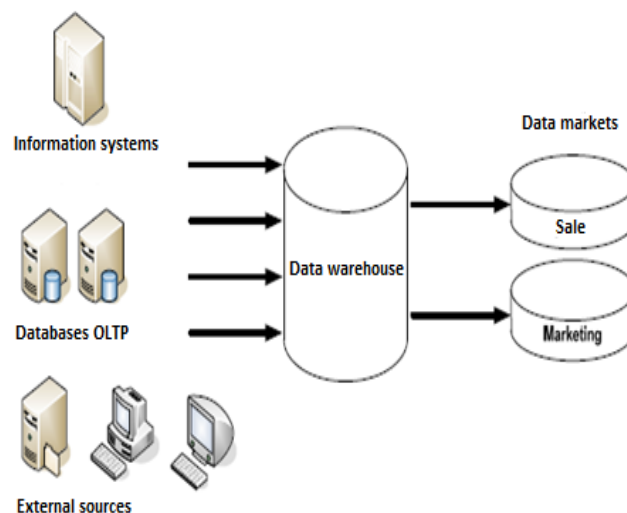


Fig. 3 Incremental method “top down”

IV. RELATIONS AND MULTIDIMENSIONAL DATABASES

Both types of databases have their advantages and disadvantages, from which then comes a scenario of deployment and use.

In a relation database data are stored in two-dimensional tables. Each row in the table contains data that is usually the image of the real world, so data that relating to any item or to it part. Columns of two-dimensional database tables contain data of the attribute. [7]

A. Multidimensional model

Simply OLAP cube is actually equivalent to a table in a relation database. From a programming point of view this is a type of multidimensional array. Each cube has several dimensions (equivalent index field in a relation database). Probably the best you can imagine a classical three-dimensional cube, but the number of dimensions in real multidimensional database is usually larger. Space for the whole cube is already scheduled in advance. [7]

Individual records are in multidimensional cubes located at the intersection of dimensions. With the increasing number of dimensions, are also rapidly growing the requirements for storage capacity. Not at all intersections of dimensions are always data present. Such cube is called a sparse cube. In practice are in multidimensional database used different technology to compress the volume of the disk space. [10]

B. Multidimensional OLAP

For multidimensional online analytical processing are data obtained either from the data warehouse or operational recourses. The mechanism of MOLAP stores analytical data in own data structure and summary. During this process are counted as many results as technically feasible and time allows. The data in the repository of MOLAP type are saved as pre-calculated field. Data and index values are stored in individual fields of a multi-dimensional database. [7]

The database is organized so, that it allows fast acquisition of the data from multiple dimensions. Part of the data can be introduced from the server to the client, which allows fast analysis without much network load.

The main advantage of MOLAP is the maximal performance in relation to user queries. Disadvantage is data redundancy, because they are stored in a relation database and also in a multidimensional database. Requirements on storage capacity can in case of multiple dimensions use, extreme increase. [7]

C. Relation databases OLAP

Relational online analytical processing (ROLAP) acquire data for the analysis from relational data warehouse. Data from the relational databases after processed, are adduced to the user as a multidimensional view.

Data and metadata are in the repository ROLAP stored as records in a relation database. OLAP server dynamically uses this metadata to generate SQL statements, which are needed to retrieve data requested by the user. In this process remain data stored in relational databases, so there is no problem with redundancy. [7]

D. Hybrid OLAP (HOLAP)

Hybrid OLAP is a combination of MOLAP and ROLAP storage, while benefiting various types of repositories, and largely eliminates the disadvantages. The data remain in relation database and counted aggregations are stored in multidimensional structures.

By consulting, the data are collected into a multidimensional memory cache. In hybrid solution relation database stores many detailed data and multidimensional model stores summary data. [7]

V. ETL (EXTRACTION, TRANSFORMATION, LOADING)

Term ETL, which stands for extraction, transformation, and loading is scheduled data integration processes. This process includes three main steps:

- extracting data from their operational or external data sources,
- transforming the data into an appropriate format,
- and loading the data into a data warehouse repository.

The goal of the first step “extraction” is to collect data from its sources. The second step “transformation” is to convert, reformat, cleanse data into format that can be used by the target database. The last step “loading” is import, because its transformed data into a target database, data warehouse, or a data mart. [6]

A. Extraction

Extraction of an ETL process involves connecting to the source systems, and both select and collect the necessary data, which are needed for analytical processing within the data warehouse. Usually data is consolidated from numerous, disparate source systems that may store the data in a different format. Thus the extraction process must convert the data into a format suitable for transformation processing. The complexity of the extraction process may vary and it depends on the type and amount of source data. [6]

B. Transformation

The transformation, as second step of an ETL process involves execution of rules to the extracted data to convert it to standard format. It also includes validation of records and their rejection if they are not acceptable. The amount of manipulation needed for transformation process depends on the data.

Good data sources will require little transformation, whereas others may require one or more transformation techniques to meet the business and technical requirements of the target database or the data warehouse. The most common processes, which are used for transformation, are conversion, clearing the duplicates, standardizing, sorting, translating and looking up or verifying if the data sources are inconsistent. [6]

C. Loading

The load is the last step of ETL process and it involves importing extracted and transformed data into a target warehouse. Some load processes physically insert each record as a new row into the table of the target warehouse. Whereas other load processes include a massive bulk insert of data utilizing a bulk load routine.

The SQL insert is a slower routine for imports of data, but does allow for integrity checking with every record. The bulk load routine may be faster for loads of large amounts of data, but does not allow for integrity check upon load of each individual record. [6]

D. Possible errors of ETL system

ETL process is not always successful, because there might occur problems with the reliability of data storage (disks are mechanical devices that wear out), can lead to loss of communication, data sources may change (by OLTP system upgrades, which are not document in the metadata).

It is important to verify the data, because if they are not verified, it may cause problems for the extraction and transmission. Based on the seriousness of the failure, it is necessary to start a new process, or it can continue from the point of the failure. Inaccurate or incomplete information may be cause of the inaccuracy results of the analysis. [6]

Because of this problems, it is necessary to test the ETL stages through simulations and then also on real data. In the testing phase will first time rebound, if the ETL is well documented. Even after testing and deploying of ETL system, we don't have full confidence that everything will always work correctly.

Data volume storage is growing rapidly and implementation and granularity metric of data requires regular review. In ETL process design can be used specialized tools from external suppliers, or developed for a specific project. (Microsoft DTS, DPS ...). [6]

VI. DESIGN OF DATA WAREHOUSE

A data warehouse is designed with use of architecture, which contains floating data repository staging area. Because of relationship between table of facts' and table of dimensions, is used snowflake architecture. The proposal of the tables is shown in Fig. 4.

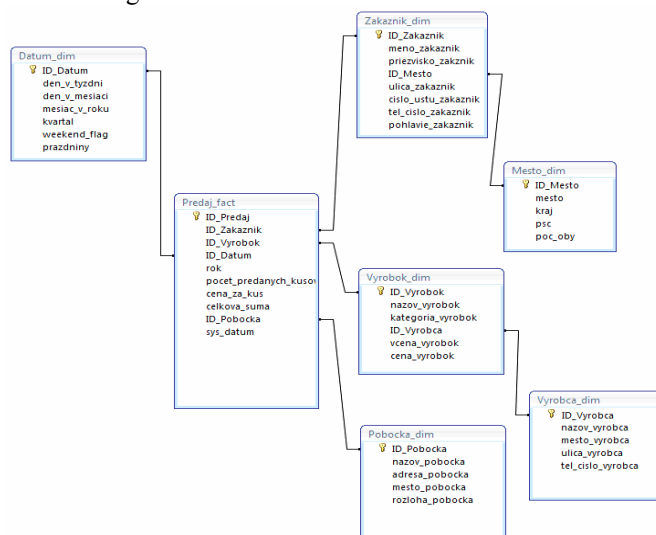


Fig. 4 Layout of data warehouse

The proposal of the individual parts is conditional on need to execute the part in specific order (Fig. 5), so that the next phase start first after the previous phase is successfully completed. Failure of this sequence would lead to incomplete data in warehouse or it could be a problem in determining of the foreign keys.

In the first phase is solved the filling of dimensions, which didn't contains foreign keys, and according to this they can be stored in warehouse at once and in random order.

In the second phase are filled dimensions containing foreign keys. The third stage performs the final factual table, which is directly dependent on the dimensions. These dimensions are the source of its foreign keys.

Use of staging area lighten input database in some operations and allows to prepare data for the introduction in a form in which they will be stored in a data warehouse. All phases are executing similar operations over the input

data and differ only in input and output tables, in the need of transformation or foreign key assignments.

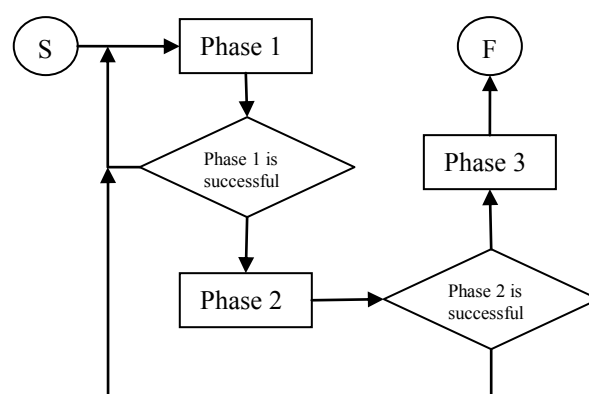


Fig. 5 Phase sequence

A. First Phase

The first phase of implementation is aimed at filling the dimensions, which don't contain any foreign keys. So they can be introduced into the data warehouse independently. These dimensions are Mesto_dim, Vyroba_dim and Pobocka_dim. After the introduction, they will be a data source to assign foreign keys to dimensions Zakaznik_dim, Vyrobock_dim, which are filled in the second phase.

At the beginning of Phase 1 is tested the connection with input databases con1.mdb, con2.mdb, database staging area stage.mdb, data warehouse database dw.mdb and presence of input flat file mesto.csv. When the database is connected, it will start with data uploading to the staging area.

Finally, connection with databases is released, because between cycle performance of data warehouse can also for hours, and it isn't necessary to have databases still opened.

B. Second Phase

The second phase link to the first phase, and starts after the first phase is successfully completed. Second phase fill data into the dimensions Zakaznik_dim, Vyrobock_dim in data warehouse stored in dw.mdb from input databases con1.mdb and con2.mdb. This phase also assigns foreign keys.

For Zakaznik_dim it is ID_mesto, which represent information about customer and is the primary key of dimension Mesto_dim.

For Vyrobock_dim it is the foreign key ID_Vyroba, which represent detailed information about the manufacture of dimension Vyroba_dim. The order of staging area fulfillment doesn't matter because dimensions Zakaznik and Vyrobock are unrelated. For this reason it doesn't matter which dimension is filled first. At the beginning is opened connection to the input database, staging area database and also with the data warehouse database.

C. Third Phase

Phase 3 will be implemented only after first and second Phase is successfully finished, because without the previous two phases wouldn't have enough data to fill factual table. In this table Predaj_fact are stored foreign keys from data warehouse dimension Datum_dim, Zakaznik_dim,

Vyrobok_di and Pobocka_dim. At the beginning is opened a connection to the database con1 and con2, database staging area and also to data warehouse.

At this point, we have all tables of the data warehouse. If don't occur an error during data recording, all records will be saved in factual table and the staging area is released for further work.

If there was an error during data recording, an error message is generated and all changes since the beginning of Phase 3 in staging area are deleted

VII. CONCLUSION

The paper deals with design of ETL system with emphasis on the sample database solution, which contained most cases of data inconsistencies occurring in corporate databases. In the simulation was used a specific input databases, from which was extracted, transformed and load data into proposed warehouse.

The limitations of this solution results from the usage of specific databases and from use of database access, with low performance. This database allows full testing of ETL system without use of installing additional database systems. Proposed ETL system is an effective and cheap solution for small companies.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0008-10. And it was also supported by the Slovak Research and Development Agency under the contract No. APVV-0073-0.

REFERENCES

- [1] V. Rainardi, "Building a Data Warehouse with examples in SQL server," by *Apress*, 1ed, 2007, ISBN-10: 1590599314.
- [2] V. Scott, "Extraction, Transformation and Load Issues and approaches," in <http://www.tdan.com/view-articles/4839>.
- [3] J. Caserta, "The Data Warehouse ETL Toolkit", published by *Wiley Publishing*, 2004, ISBN 0-764-56757-8
- [4] L.Lacko, "Databáze: dative sklady, OLAP a dolování dat s příklady v MSCL Serveru a Oracle," in *Computer Press*, Brno, 2003, ISBN 80-7226-969-0.
- [5] J. Mundy, W. Thornthwaite, R. Kimball, "The Microsoft Data Warehouse Toolkit", published by *Wiley Publishing*, 2011, ISBN 978-0-470-64038-8.
- [6] G.K.Gupta, "Introduction to Data Mining with case studies", by PHI learning private limited, 2011, ISBN 978-81-203-4326-9.
- [7] Ch. S. Jensen, T.B. Pedersen, Ch. Thomsen, "Multidimensional databases and data warehousing", by Morgan and Claypool publishers, 2010, ISBN 9781608455379.
- [8] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy, B. Becker, "The data warehouse lifecycle toolkit", by Viley Publishing, Inc., ISBN:978-0-470-14977-5
- [9] W.H.Inmon, "Building the data warehouse", John Wiley & Sons, Inc. New York, 1996, ISBN:0-471-14161-5
- [10] S.Sumathi, S.N. Sivanandam, "Introduction to data mining and its applications", by Springer, 2006, ISBN-10 3-540-34350-4

Extract - Transform - Load Data Warehouse System

Anton BALÁŽ

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

anton.balaz@tuke.sk

Abstract—Information are critical for business and corporations' market survival. Data warehousing is one of the technologies which provide access to an information by variety of approaches. In this paper data warehousing is introduced with focus to design own ETL system with running processes in the data warehouse.

Keywords—data warehouse, ETL, extraction, transformation, loading, data pump

I. INTRODUCTION

Progress in engineering, financial or business sector have fetched along the accumulation of large data volumes of various formats and importance. The ability of target users to understand and use this data tend to descend. Thus, there grows a need for data analysis systems capable of automatic data analysis and summarization to provide information and predict future trends [1].

Data warehousing is the technology allowing to reach this goals. Data warehouse is not only just one software or hardware product but it represents complete computing environment where users are in touch with data they need to acquire information from [2]. Data pump and ETL system are one of the crucial parts of this environment. Therefore the challenge lies here how to design such a system.

The topic of this work is to introduce data warehousing with focus on the ETL system, its functionality, its issues and last but not least to explore current ETL systems situation and finally to try to design working ETL system based on this knowledge.

The goal of this work is to design and an ETL tool to provide basic ETL system functionality. Its main functions include data extraction from variety of sources, data transformation and loading, all user controlled.

II. ETL SYSTEM OF DATA WAREHOUSE

A. Data Warehousing

Data warehousing is a set of technologies and algorithms that gather various data from multiple sources to a single repository and provide the ability to analyze them [1]. This analysis is way too general. According to father of data warehousing Bill Inmon, data warehouse is subject-oriented, integrated, nonvolatile and time variant collection of data in support of management's decisions [2]. Let's take a closer look to understand mentioned characteristics:

- Subject-oriented - data in data warehouse are organized around main subjects (customer, sales etc.) and data warehouse focus on modeling them to support decision making, not to support everyday transactions [1].

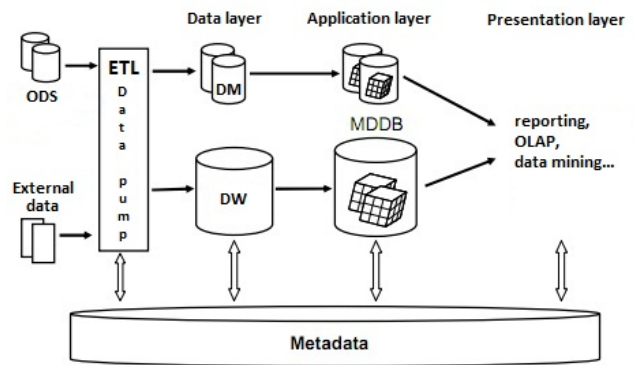


Fig. 1: Data warehouse structure

- Integrated - data warehouse is constructed by integration of multiple heterogeneous source systems in range from OLTP database systems, flat files or even web sources [1].
- Nonvolatile - Data approach for end users in data warehouse tend to be read only. Data already stored in data warehouse are not deleted or updated. Data are loaded into data warehouse in bigger loads (usually periodical) and they are kept to provide historical view for data analysis [3].
- Time variant - data in data warehouse are stored to provide information from historical perspective [3].

Data warehouse is computing environment that [2]:

- provides complete and integrated view on enterprise
- makes actual and historical enterprise information easily available to support decision making
- performs decision support transaction without putting load on operating system
- presents flexible and interactive source of strategic information

Before we focus on ETL system, we should focus what place it takes in data warehouse system and what its purpose is. Fig. 1 shows a basic layout and structure of data warehouse system.

Let's look at the structure from left to right. First level of data warehouse system is represented by data sources. We can put there various external sources (external databases, flat files, web etc.) or operational data stores (ODS). ODS is a hybrid store for unified data originated in data source systems providing comprehensive views of data in operational systems. The difference between ODS and classical relational transaction oriented database is that data are subject oriented [4]. Source data need to be transferred from their source systems to the data warehouse. This task is performed by

data pump performing ETL phase of DW system. Data pump extracts data, transforms it and load to the specified target systems - enterprise wide data warehouse (DW) or data marts (DM). We can ask if there is some difference. In some sources the concepts are used interchangeably. However, data marts are considered to be subsets of enterprise wide data warehouse and they are designated to satisfy corporation departments' specific needs. They can prove data warehouse building propriety [2]. Data warehouse and data marts are usually constructed in form of multidimensional databases (MDDB) also known as data cubes (or hypercubes). Cube is database type where data are stored in the cells. Each cell is defined by determined number of variables - dimensions. Each cell contains specific event and its dimensions define conditions of this event. This data structures provide great resource for data analysis and information aquisition. Analysis is performed by variety of tools or approaches including simple queries, reporting or more complex technologies such as OLAP or data minig. Every part of data warehouse system may be organized by metadata containg various information about data structure, importance, using etc. [5].

B. ETL System

Simple definition states that it is a mechanism to transport data from source systems to the data warehouse. Soundes easy but in fact it is one of the most complex components of data warehouse system. It should be able to work with various sources, not dedicated for single systems type [6]. ETL system is responsible for transporting data into data warehouse so it influences information gathered from them in a large extent. ETL system perform three main phases [8]:

- 1) Extraction - includes all processes necessary to connect to various data sources and to select intended data for consequent phases. Exctraction should be hardware and software independent [6].
- 2) Transformation - represents every operation performed upon extracted data. Transformation phase may include [9]:
 - Integration
 - Aggregation
 - Cleansing
 - Denormalization
 - Derivation
 - Validation
 - Summarization
- 3) Loading - last step of ETL where transformed data are loaded into data warehouse dimensional structures.

ETL system is subject to many issues we will discuss in the course of designing new ETL system in the next chapter.

III. DESIGNED ETL DATA WAREHOUSE

Main purpose of designed ETL system is its modularity, providing easy form of adding new features and functionality. Other system architecture requirements:

- setting ETL system configuration to determine extraction, transformation and loading properties
- starting ETL system according to set configuration
- saving configuration of ETL system and its loading for later use
- previewing data

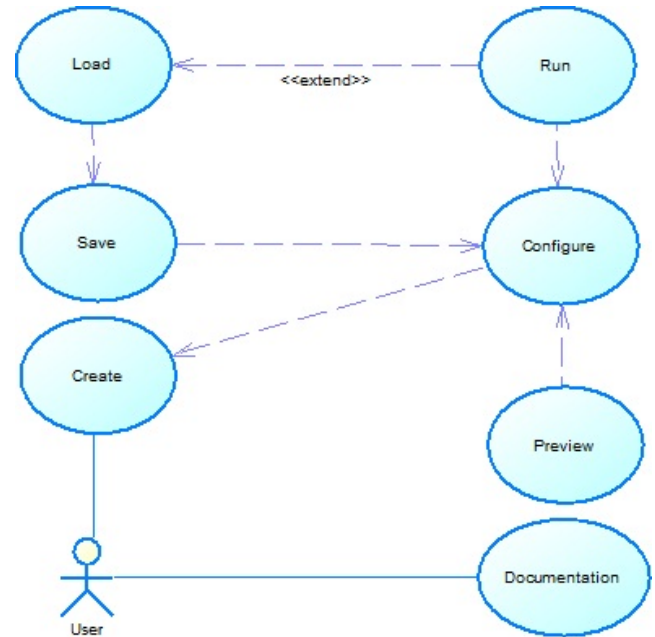


Fig. 2: ETL system function diagram

- interactive GUI

ETL system will be designed using object-oriented approach because of its suitability for this purpose. ETL system properties will be set via GUI interactively by user. User's task will be to choose from supported source systems, define target system and set transformations. Every component is graphically represented by corresponding GUI component. Chosen components need to be connected to set where will be data taken from, how they will be transformed and where they will be stored. Basic functionality can be depicted by UML use case diagram in Fig. 2 and is described in Table. I.

TABLE I: ETL tool functionality

Type	Name	Description
Actor	User	ETL tool user
Use case	Create	create ETL tool configuration using GUI
Use case	Configure	set ETL rules for created configuration
Use case	Run	run ETL
Use case	Save	save set configuration
Use case	Load	load saved configuration
Use case	Preview	preview current data in systems acting in ETL
Use case	Documentation	find help in user or system manuals

Main part of the system functionality is to start system to extract, transform and load data according to its configuration. Multiple issues must be treated.

- First issue concerns user privileges on used database sources or targets. User must ensure he is granted privileges allowing him to perform demanded operations. The privileges include privilege to retrieve (select) data from data sources or write (insert/update) data into target systems. Target systems must be also unlocked to allow inserting or updating data.
- Next issue concerns primary key control. The system must provide such mechanism of data loading that will not violate primary key constraints in target databases. The solution for this issue can be primary key generator

for generating new primary keys, generating common primary keys or generating surrogate primary keys. In designed ETL tool, new primary key generator will be used. It generates random primary key bounded by datatype and maximum length of target data warehouse table attribute.

- Datatypes needs to ensure data integrity. Source attributes datatypes and target attributes datatypes must be equivalent or implicitly convertible. In the new system, datatypes conversion will be controled to prevent mismatches. Since there are sources of multiple types in ETL systems, every type uses its own datatypes. Therefore it is necessary to find a way how to transport data of source systems' datatypes to the target systems satisfying its datatypes' rules and conditions. In our case we will convert each datatype to the programming language standard datatypes in extraction phase and convert them back in loading phase.
- Another data issue is data duplicity. Data in data warehouse should be duplicity as little as possible because information gathered could be misrepresented. In our new ETL system we will target this issue with two methods. First method installs using of two different loading modes - INSERT and UPDATE mode. In case of INSERT mode, while inserting data into data warehouse, if there already is a record in data warehouse with the same primary key value as inserted record, new primary key is generated (see primary key control above) and new record is inserted into data warehouse no matter if other record values are equal. This allows some duplicity. In case of UPDATE mode, if there is a record found in data warehouse with same primary key value, other record values are updated to values of new record being inserted. Second method allows user to clear all data from data warehouse before transformation and load processes and to start its new refill.
- Data issues also include missing values issues or NULL values issue. In that case data augmentation or derivation is required. We can use default values for missing or NULL values to be inserted into data warehouse tables providing correct format of default values. Another concept use metadata or other available data to derive new data to be inserted.
- Last major issue we will face is referential integrity issue. Again some derivation or default values selected from parent table may be used.

Structure of the system is depicted by schema in Fig. 3. Logic of the system and its GUI will be separated. Logic consists of supported systems' classes (in our case database systems Oracle, MS SQL Server and MySQL, CSV and XML files) and supported transformation classes. Each transformation class represent the way of loading extracted data into target structure. Transformation of data is provided by functions included inside transformation class. There are also primary key generator class and type converter class to face formerly mentioned ETL issues. GUI part of the system consists of various forms and components to represent created objects. Components will be placed inside forms by user and they will be interconnected the same way to define how ETL will be executed. Connection between logic and GUI will be provided by representation of each created object by corresponding GUI

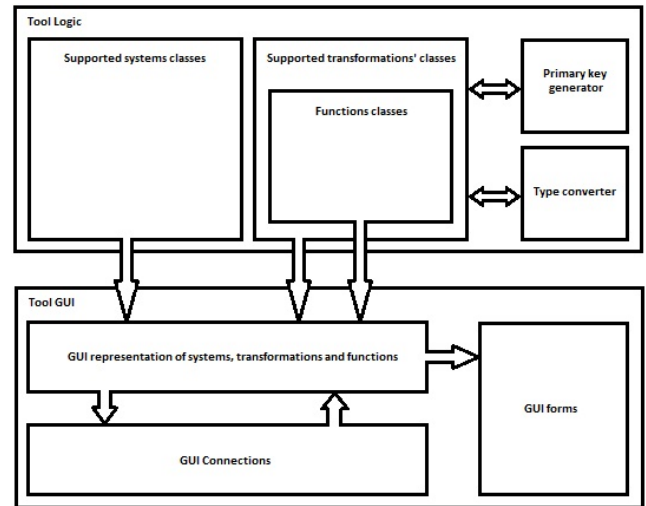


Fig. 3: ETL system structure schema

object. Structure should be modular as much as possible. Basic modularity is provided by use of object-oriented approach itself but more is required. Each supported ETL source system and transformation will be defined in individual class so the tool can be easily augmented by creating new classes containing new source system or operation definitions.

Complete functional and system structure are introduced in the main part of the system in form of UML use case and class diagrams. Following created system design and treating the issues we can step into implementation of the ETL tool. .NET Framework and object-oriented programming language C# are suitable for this purpose.

IV. CONCLUSION

The goal of this work was to design ETL tool capable of performing simulation of data warehouse ETL system. The tool was tested and practical simulation proved it is capable of extracting, transforming and loading data into chosen target systems.

The reason of designing and implementing of the system was the aim to point at its necessity in data warehousing, its overall complexity and to bring own solution which can be used for personal use not only with data warehouse but also with classical types of databases. It can also serve as data migrator - transferring data from source systems to target systems.

It was designed for object-oriented implementation with its own GUI. This approach brought some advantages. Our ETL tool is modular and easily extensible and it is simple to maintain and modify its components.

ETL tool allows user to configure ETL system interactively. It also allow periodical data loading via saving and loading created system configurations. Anytime data can be viewed to check their transformation and target systems content changes.

The tool has some disadvantages and limitations. It is dedicated to simulate processes of data warehouse ETL system, not to replace it completely. It lacks complex metadata system so complex operations such as derivation or validation are missing.

ACKNOWLEDGMENT

This paper was developed with support of the Project "Centrum excelentnosti integrovaného výskumu a využitia

progresívnych materiálov a technológií v oblasti automobilovej elektroniky” (Centre of Excellence of Integrated Research and Exploitation the Advanced materials and Technologies in the Automotive Electronics), ITMS 26220120055, that is co-financed from Structural Funds EU ERDF within Operational Program Research and Development OPVaV-2009/2.1/03-SORO and preferred axis 2 Support of Research and Development.



REFERENCES

- [1] Singhal Anoop: Data warehousing and data mining techniques for cyber security. 13 Dec. 2006, 172 p. ISBN 0387264094.
- [2] Ponniah Paulraj: Data warehousing fundamentals: A comprehensive guide for IT professionals. 2001. 544 p. ISBN 0471412546.
- [3] Paralič Ján: Objavovanie znalostí: Prednášky: Dátové sklady. FEI TU Košice. Košice 2010.
- [4] Humphries Marc, Hawkins Michael W., Dy Michelle C.: Data warehousing: Architecture and implementations. 1999, 360 p. ISBN 0130809020.
- [5] Rainardi Vincent: Building a Data Warehouse: With examples in SQL Server. 2007. 523 p. ISBN 1590599314.
- [6] Casters Matt, Bouman Roland, Van Dongen Jos: Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration. 2010. 720 p. ISBN 9700470635719.
- [7] Ádám N. - Single input operators of the DF KPI system. In: Acta Polytechnica Hungarica. Vol. 7, no. 1 (2010), p. 73-86. - ISSN 1785-8860
- [8] Kimball Ralph, Caserta Joe: The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. 2004. 491 p. ISBN: 0764579231.
- [9] Tanuška, Pavol: Vývoj informačných systémov : Prednášky : Dátové sklady Data Warehousing. KAIA MTF STU Bratislava. Trnava 2006.
- [10] Látka O., Madoš B., Perháč J., Kleinová Alžbeta Parallel system module for prepare photorealistic rendering in grid and cluster SAMI 2008 Proceedings, 6th International Symposium on Applied Machine Intelligence and Informatics, Košice - Herľany, 21.-22.1.2008, 2008, 6, pp. 317-320, ISBN 978-1-4244-2106-0

Filter Bank Multicarrier Transmission System

Denis DUPÁK (2st year), Radovan BLICHA (3rd year)
Supervisor: Dušan KOCUR

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

denis.dupak@tuke.sk, radovan.blicha@tuke.sk, dusan.kocur@tuke.sk

Abstract— Orthogonal frequency division multiplexing (OFDM) is the most favored modulation nowadays for high speed digital communications, but many other modulation techniques show advantages over it. The major disadvantage of OFDM signal is its large envelope fluctuation, which significantly limits using of high power amplifiers (HPA) whose transmission characteristics exhibit a nonlinear character. In this paper we present the Filter bank multicarrier (FBMC) transmission system which is one of the possible solutions to reduce a nonlinear distortion. We describe the main principle of FBMC and compare this method to OFDM via bit error rate (BER) system performance.

Keywords—Orthogonal Frequency Division Multiplexing (OFDM), Filter Bank Multicarrier (FBMC), Offset QAM (OQAM).

I. INTRODUCTION

The demand for high data rates transmission over multipath radio channels has increased rapidly. To fulfill this requirement, utilization of multi-carrier based transmission techniques seems to be an inevitable solution. Very promising modulation technique which satisfy communication system requirements is OFDM. The OFDM is spectrally efficient modulation technique that has recently gained much popularity due to high speed and very effective data transmission over multipath fading channels. The OFDM based transmission systems are characterized by large number of benefits in comparison with traditional schemes. This makes it superior compared to previously introduced wireless standards [1].

On the other hand, OFDM has some well known drawbacks. To perform interference free demodulation the subcarriers have to remain orthogonal, otherwise system performance will degrade. Therefore, OFDM is very sensitive, to frequency offset originating from the mismatch of transmitter and receiver local oscillators. The time domain OFDM signal is a sum of a large number of complex sinusoids, which means that, according to the central limit theorem, the amplitude distribution will be Gaussian, leading to a large peak to average power ratio (PAPR) of the signal. Hence, a power amplifier with a relatively large linear range is required, otherwise non-linear effects will severely degrade the system performance [2].

Large sensitivity to nonlinear amplification has greatly limited the practical applications of OFDM transmission systems. In order to alleviate the effect of nonlinear

amplification in OFDM transmission systems, many approaches based on different techniques have been introduced. Frequently used solution in the transmitter is to back-off the operating point of nonlinear amplifier, but this approach results in significant power efficiency penalty. Alternative approaches for OFDM performance improvement are realized by applications of other usually computational demanded PAPR reduction methods at the transmitter side, for example pre-coding of signals (Single Carrier Frequency Division Multipl Access-SC-FDMA), active constellation extension [3], tone reservation [4] or selected mapping [5]. Another well known and promising solutions to reduce influence of nonlinear distortion are the application of nonlinear iterative detection [6] at the receiver side.

Another design solution of multi carrier transmission system based of application of bank of filters is FBMC transmission system which is more resistant to nonlinear distortion than OFDM transmission systems. In this article, the basic principle of FBMC transmission systems is presented.

II. HISTORICAL OVERVIEW OF FBMC

The first proposal came from Chang [7], who presented the conditions required for signaling a parallel set of pulse amplitude modulated (PAM) symbol sequences through a bank of overlapping filters within a minimum bandwidth. To transmit PAM symbols in a bandwidth efficient manner, Chang's signaling is based on staggering a number of overlapping vestigial side-band (VSB) modulated signal sequences. Saltzberg [8], extended the idea and showed how the Chang's method could be modified for transmission of quadrature amplitude modulated (QAM) symbols, in a double side-band modulated format. Efficient digital implementation of Saltzberg's multicarrier system through polyphase structures was first introduced by Bellanger [9] and later studied by Hirosaki [10]. Another key development appeared in [11], where the authors noted that Chang's/Saltzberg's method could be adopted to match channel variations in doubly dispersive channels, hence, minimize intersymbol interference (ISI) and interchannel interference (ICI). This point, which is extensively elaborated on in this article, can lead to a significant improvement of FBMC over OFDM in highly mobile channels. Both Chang's and Saltzberg's methods belong to a class of multicarrier techniques that may be referred to as FBMC systems.

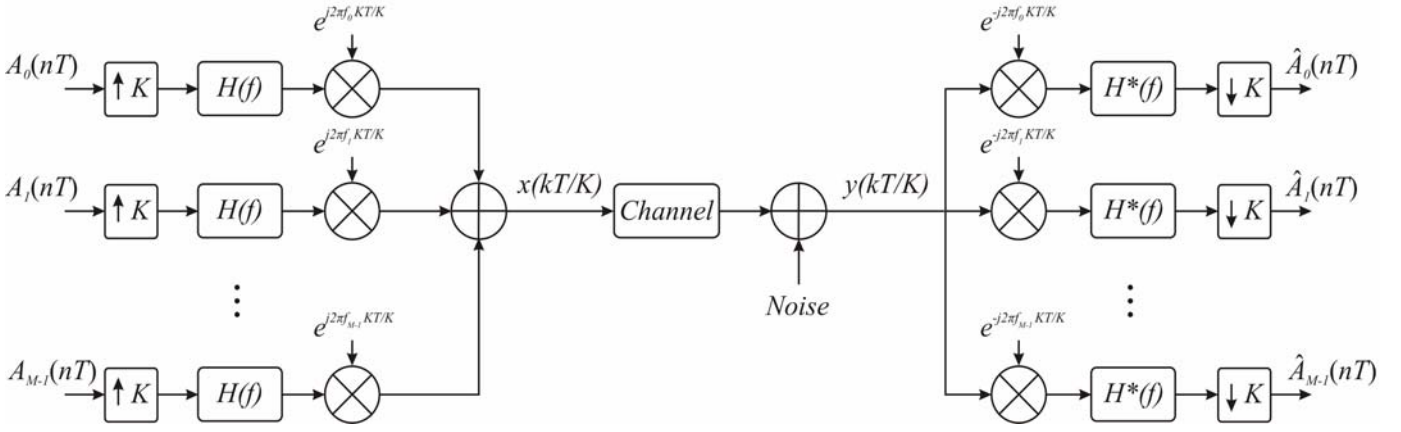


Fig. 1. Block scheme of FBMC transmission system

III. FBMC TRANSMISSION SYSTEM

The idea of subdividing a signal frequency band into a set of contiguous bands in order to achieve efficient system realizations has pervaded the fields of signal processing and data communications for many decades. Filter banks [12] have emerged as a powerful analysis and design technique in that context.

Fig. 1 shows an M -subchannel filter-bank multi carrier transmission system. The complex-valued modulation symbols $A_m(nT)$, $m=0, 1, \dots, M-1$, chosen from not necessarily identical QAM constellations, are provided at the symbol rate of $1/T$. After upsampling by a factor of K , indicated by the notation $\uparrow K$, each symbol stream is filtered by a baseband filter, referred to as a prototype filter, with frequency characteristic $H(e^{j2\pi f})$ and impulse response $h(k)$. The transmit signal $x(kT/K)$ is obtained at the transmission rate of K/T by adding the M filter-output signals properly shifted in frequency. At the receiver, matched filtering (where $*$ denotes complex conjugation) is employed, followed by subsampling by a factor of indicated by the notation $\downarrow K$. When, the $K=M$ ($K>M$)

The filter bank is said to be critically (noncritically) sampled. Fig. 2 illustrates the way in which the parameters K , M , and $1/T$ determine the spectral characteristic of the transmitted signal.

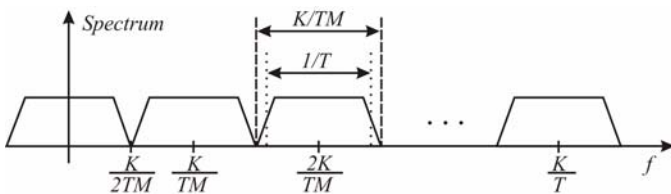


Fig. 2. Spectral characteristic of transmitted FBMC signal

The inherent signal shaping capability of filter-bank systems offers several degrees of freedom to the communication system designer. Traditionally, filter characteristics have been chosen to satisfy a “perfect reconstruction” constraint in order to ensure that transmission is free of ISI within a subchannel as well as free of ICI.

We can express the transmitted FBMC signal by this formula:

$$x\left(k \frac{T}{K}\right) = \sum_{m=0}^{M-1} \sum_{n=-\infty}^{+\infty} A_m(nT) h\left[\left(k-nK\right) \frac{T}{K}\right] \times e^{j2\pi m(K/MT)k(T/K)}. \quad (1)$$

The difference between OFDM and FBMC lies in the choice of T and the transmitter and receiver prototype filters. In OFDM systems is used rectangular pulse of height one and width T . The receiver prototype filter is also a rectangular pulse of height one, but its width is reduced to $T_{FFT} < T$, where $T_{FFT} = 1/B$ and B is the frequency spacing between subcarriers. In FBMC systems that are designed for maximum bandwidth efficiency, $T = T_{FFT} = 1/B$, however, the durations of pulses are greater than T (usually, an integer multiple of T). Hence, in FBMC, the successive data symbols overlap.

Major problem occurs in FBMC transmission systems by the transmission of complex modulation symbols. Imaginary and real part of the modulation symbol have to be transmitted separately in different sub-channels, thus the data rate of the transmitted system is significantly reduced. This problem is solved by applying the offset-QAM (OQAM) [13]. The basic principle of OQAM modulation is that the real and imaginary component of the data symbols are transmitted separately on two different sub-carrier signals which are shifted by $K/2TM$. Spectral characteristic of OQAM FBMC signal is illustrated in Fig.3.

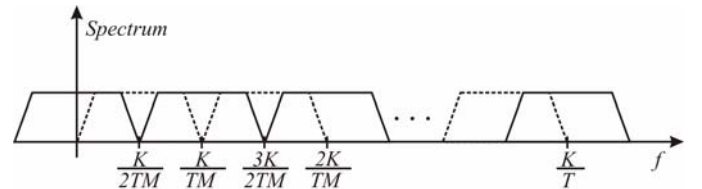


Fig. 3. Spectral characteristic of OQAM FBMC signal

It is obvious that number of subcarrier signals are doubled. Imaginary parts of complex data symbol are transmitted on subcarrier signals on frequencies $(2m+1)K/2TM$ and real parts are transmitted on frequencies mK/TM . We can see on Fig.3 that spectra of FBMC symbols are overlapped in frequency domain. To prevent a rapid increase in the ICI due to overlapping spectrum FBMC symbols, the orthogonal functions sine and cosine are used in OQAM modulation [14]. By applying this modulation method there are no loss of data rate. With the use of OQAM modulation, no data rate loss

will occur. The other advantage of FBMC is that no cyclic prefix (CP) has to be used to combat the channel-induced intersymbol interference. Conversely a more complex signal processing has to be applied, and the channel equalization in the receiver chain will be more complex than other schemes.

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.

In the present work, we have used QAM modulation for OFDM transmission system and OQAM for FBMC transmission system. We have compared those systems in linear and nonlinear channel. The soft limit model of HPA was used. Both transmission system used 256 subcarrier signals.

In Fig.4 is illustrated BER performance of FBMC and OFDM transmission systems when 16-QAM was used, where E_b is transmitted energy per bit and N_0 is power spectral density of Additive white Gaussian noise (AWGN). We can observe that FBMC system outperform OFDM system in linear channel and also in nonlinear channel. Performance gain is equal to 3dB.

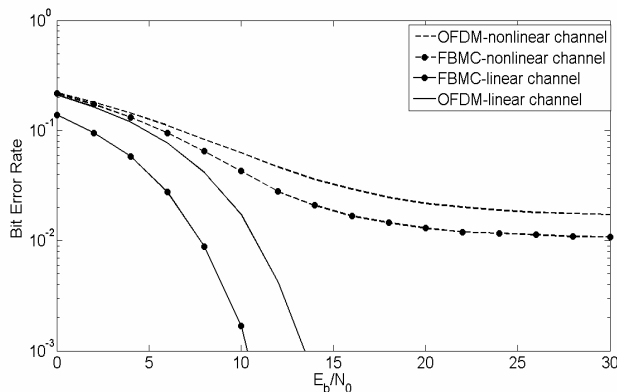


Fig. 4. Bit error rate of OFDM and FBMC transmission systems when 16-QAM was used

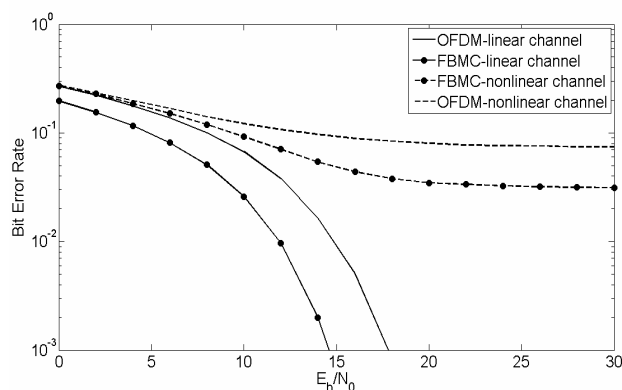


Fig. 5. Bit error rate of OFDM and FBMC transmission systems when 64-QAM was used

In Fig.5 is illustrated BER performance of FBMC and OFDM transmission systems when 64-QAM was used. We can observe that FBMC system outperform OFDM system in linear channel and also in nonlinear channel. Performance gain is equal to 3dB.

V. CONCLUSION

In this paper, we have presented basic principle of FBMC transmission systems. Those systems are characterised with better features in comparison with OFDM systems in nonlinear and also linear channel. As it can be seen in previous section, the performance gain in terms of BER is about 3dB. Performance gain can be increased by application of additional methods which were used in OFDM systems for example iterative receiver, multidimensional baseband modulation, etc. In future we will focus on implementation of those methods in FBMC transmission systems.

ACKNOWLEDGMENT

This work was supported by the Slovak Cultural and Educational Grant Agency (KEGA) under the contract No. 010TUKE-4/2012.

REFERENCES

- [1] E. Dahlman, S. Parkvall, "3G Evolution, HSPA and LTE for Mobile Broadband. Academic Press", Oxford UK, 2007.
- [2] H. G. Myung, D. J. Goodman, "A New Air Interface for Long Term Evolution.", Winchester, UK: Wiley, 2008.
- [3] B. S. Krongold, D. L. Jones, "PAR Reduction in OFDM via Active Constellation Extension", *IEEE Transactions on Broadcasting*, vol. 49, no. 3, pp. 258-268, Sep. 2003.
- [4] M. Duemal, A. Behravan, T. Eriksson, J. L. Pijoan, "Evaluation of performance improvement capabilities of PAPR reducing methods", *Wireless Personal Communications*, vol. 47, No. 1, pp. 137-147, Oct. 2008.
- [5] L. J. Cimini, N. R. Sollenberger, "Peak-to-Average Power Ratio Reduction of an OFDM Signal Using Partial Transmit Sequences", *IEEE Communications Letters*, vol. 4, No. 3, pp. 86-88, Mar. 2000.
- [6] J. Gazda, "Multicarrier based transmission system undergoing nonlinear amplification", Doctoral Thesis, Košice: Technical University of Košice, Faculty of Electrical Engineering and Informatics, 2010.
- [7] R. W. Chang, "High-speed multichannel data transmission with bandlimited orthogonal signals," *The Bell System Technical Journal*, vol. 45, pp. 1775-1796, 1966.
- [8] B. R. Saltzberg, "Performance of an efficient parallel data transmission system," *IEEE Transactions on Communications Technology*, vol. 15, no. 6, pp. 805-811, 1967.
- [9] M. Bellanger and J. Daguette, "TDM-FDM transmultiplexer: Digital polyphase and FFT," *IEEE Trans. Commun.*, vol. 22, no. 9, pp. 1199-1205, Sept. 1974.
- [10] B. Hirotsu, "A n orthogonally multiplexed QAM system using the discrete Fourier transform," *IEEE Trans. Commun.*, vol. 29, no. 7, pp. 982-989, July 1981.
- [11] B. Le Floch, M. Alard, and C. Berrou, "Coded orthogonal frequency division multiplex," *Proc. IEEE*, vol. 83, no. 6, pp. 982-996, June 1995.
- [12] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
- [13] M. Bellec and P. Pirat. (2006, Jan.). OQAM performances and complexity. IEEE P802.22 Wireless Regional Area Network [Online]. Available: http://www.ieee802.org/22/Meeting_documents/2006_Jan/index.html
- [14] T. Fusco and M. Tanda, "Blind frequency-offset estimation for OFDM/OQAM systems," *IEEE Trans. Signal Processing*, vol. 55, no. 5, pp. 1828-1836, May 2007.
- [15] H. S. Sourck, Y. Wu, J. W. M. Bergmans, S. Sadri, and B. Farhang-Boroujeny, "Complexity and performance comparison of filter bank multicarrier and OFDM in uplink of multicarrier multiple access networks," *IEEE Trans. Signal Processing*, accepted for publication.

Fuzzy control application for a class of nonlinear systems

Vladimír SERBÁK (1st year)

Supervisor: Anna FILASOVÁ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

vladimir.serbak@tuke.sk, anna.filasova@tuke.sk

Abstract—The paper presents the stabilizing controller application for a class of continuous-time nonlinear systems represented by Takagi-Sugeno-Kang models, and measurable premise variables. Based on an enhanced Lyapunov inequality, the conditions are outlined in the terms of linear matrix inequalities to possess a stable structure closest to optimal asymptotic properties. Simulation results illustrate the design procedure and demonstrate the basic performances of the proposed control design method.

Index Terms—Takagi-Sugeno-Kang models, fuzzy control, linear matrix inequalities, convex optimization.

I. INTRODUCTION

For nonlinear system design, various control schemes were introduced including exact feedback linearization and adaptive control. The technique of exact feedback linearization needs perfect knowledge of the nonlinear system and uses that knowledge to cancel the nonlinearities of the system. Since perfect knowledge of the system is almost impossible, the technique of exact feedback linearization cannot be generally used for nonlinear system control design. Also adaptive control schemes, which were introduced to deal with nonlinear systems, exploit complicated parameter update laws and so the adaptive control algorithms possess hard limitations.

Since a generic method for design of a controller valid for all types of nonlinear systems has not been developed yet, an alternative to design a controller for nonlinear systems is e.g. fuzzy approach which benefits from the advantages of the approximation techniques approximating nonlinear system model equations. Using the Takagi-Sugeno-Kang (TSK) fuzzy model [12] the nonlinear system is represented as a collection of the fuzzy rules, where each rule utilizes the local dynamics by a linear system model. Since TSK fuzzy models can well approximate a large class of nonlinear systems, and the TSK model based approach can apprehend the nonlinear behavior of a system while keeping the simplicity of the linear models, by employing the TSK fuzzy model a control design methodology exploits fully advantage of the modern control theory, especially in the state space optimal and robust control.

The main idea of the TSK model-based controller design is to derive control rules so as to compensate each rule of a fuzzy system, determining the local feedback gains [10],

[11]. It is known that the separate stabilization of these local models does not ensure the stability of the overall fuzzy model, and global design conditions have to be used to guarantee the global stability and control performance. Therefore, a range of stability conditions have been developed for TSK fuzzy systems [1], [3], [14], most of them relying on the feasibility of an associated system of linear matrix inequalities (LMI) [2]. Therefore, the state control based on fuzzy TSK systems model gives control structures which can be designed using technique also based on equivalent LMIs (some principles and results are reported e.g. in [4], [8], [13]). The idea behind this type of design is that the TSK model based fuzzy control provides a user-friendly formalism, [16] for representing, implementing and achieving high-performance control structures.

The main contribution of the paper is to present a the stabilizing fuzzy state control application for continuous-time nonlinear MIMO systems approximated by a TSK model, and using the measurable premise variables and to compare the efficiency with the LQ controller.

II. ON THE TAKAGI-SUGENO-KANG FUZZY MODEL

The systems under consideration is one class of multi-input and multi-output nonlinear (MIMO) dynamic systems, represented in state-space form as

$$\dot{\mathbf{q}}(t) = \mathbf{a}(\mathbf{q}(t)) + \mathbf{B}\mathbf{u}(t) \quad (1)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{q}(t) \quad (2)$$

where $\mathbf{q}(t) \in \mathbb{R}^n$, $\mathbf{u}(t) \in \mathbb{R}^r$, $\mathbf{y}(t) \in \mathbb{R}^m$, are vectors of the state, input, and output variables, respectively, and $\mathbf{B} \in \mathbb{R}^{n \times r}$, and $\mathbf{C} \in \mathbb{R}^{m \times n}$ are real finite values matrices.

It is assumed that $\mathbf{a}(\mathbf{q}(t))$ is a vector function of the vector variable $\mathbf{q}(t)$ only, is bounded in associated sectors, i.e. in the fuzzy regions where it is assumed the system will operate within, and takes the value $\mathbf{a}(0) = 0$. It is considered that the number of the nonlinear terms in the nonlinear part of model $\mathbf{a}(\mathbf{q}(t))$ is p , and that there exists a set of nonlinear sector

functions of this properties

$$\begin{cases} w_{lj}(\theta(t)), j = 1, 2, \dots, k, l = 1, 2, \dots, p \\ w_{l1}(\theta(t)) = 1 - \sum_{j=2}^k w_{lj}(\theta(t)) \\ w_{lj}(\theta(t)) = w_{lj}(\theta_j(t)) \end{cases} \quad (3)$$

where k is the number of sector functions, and

$$\theta(t) = [\theta_1(t) \quad \theta_2(t) \quad \dots \quad \theta_q(t)] \quad (4)$$

is the vector of premise variables. It is supposed in the next that premise variables are measurable, and a premise variable can represent the state variable.

Using a TSK model, the conclusion part of a single rule consists no longer of a fuzzy set [9], but determines a function with state variables as arguments, and the corresponding function is a local function for the fuzzy region that is described by the premise part of the rule. Thus, using linear functions, a system state is described locally (in fuzzy regions) by linear models, and at the boundaries between regions a suitable interpolation is used between the corresponding local models.

Thus, the aggregated function set $\{w_i(\theta(t)), i = 1, 2, \dots, s, s = 2^k\}$ can be constructed from all combinations of the sector functions which implies

$$\dot{q}(t) = \sum_{i=1}^s h_i(\theta(t)) \Omega_i(t), \quad h_i(\theta(t)) = \frac{w_i(\theta(t))}{\sum_{i=1}^s w_i(\theta(t))} \quad (5)$$

$$\Omega_i = A_i q(t) + B u(t) \quad (6)$$

where $h_i(\theta(t))$ is the i -th aggregated normalized membership function satisfying conditions

$$0 \leq h_i(\theta(t)) \leq 1, \quad \sum_{i=1}^s h_i(\theta(t)) = 1 \quad \forall i \in \{1, \dots, s\} \quad (7)$$

Therefore, the fuzzy approximation of (1) leads to (5), (6) where $A_i \in \mathbb{R}^{n \times n}$ is the Jacobian matrix of $a(q(t))$ with respect to $q(t) = q_i$, and q_i is the center of the i -th fuzzy region described by the set of sector functions (3).

Assumption 1: The matrices B, C are the same for all local models.

Assumption 2: The pair $(A(q(t)), B)$ is locally controllable and $B (C)$ is of full column (row) rank, where

$$A(q(t)) = \sum_{i=1}^s h_i(\theta(t)) A_i \quad (8)$$

Now the TS fuzzy model for (1), (2) takes form

$$\dot{q}(t) = \sum_{i=1}^s h_i(\theta(t)) (A_i q(t) + B u(t)) \quad (9)$$

$$y(t) = C q(t) \quad (10)$$

and can be interpreted as a combination of s linear sub-models through the set of normalized membership functions $\{h_i(\theta(t)), i = 1, 2, \dots, s\}$ to approximate the nonlinear system.

III. STABILIZING FUZZY CONTROLLER

A. Standard Fuzzy Control Design

Considering the generalized form of (9)

$$\dot{q}(t) = \sum_{i=1}^s h_i(\theta(t)) (A_i q(t) + B u(t)) \quad (11)$$

and using the same set of membership function, the nonlinear fuzzy state controller is defined as

$$u(t) = - \sum_{j=1}^s h_j(\theta(t)) K_j q(t) \quad (12)$$

Substituting (12) into (11) results in

$$\begin{aligned} \dot{q}(t) &= \\ &= \sum_{i=1}^s h_i(\theta(t)) (A_i q(t) - \sum_{j=1}^s h_j(\theta(t)) B K_j q(t)) \end{aligned} \quad (13)$$

Since $\sum_{i=1}^s h_i(\theta(t)) = 1 \quad \forall i \in \{1, \dots, s\}$ it yields

$$\dot{q}(t) = \sum_{i=1}^s \sum_{j=1}^s h_i(\theta(t)) h_j(\theta(t)) (A_i - B K_j) q(t) \quad (14)$$

The equilibrium of the fuzzy system (9), (10) controlled by the fuzzy controller (12) is globally asymptotically stable [7] if there exists a positive definite matrix $X \in \mathbb{R}^{n \times n}$ and matrices $Y_j \in \mathbb{R}^{r \times n}$ such that

$$X = X^T > 0 \quad (15)$$

$$X A_i^T + A_i X - Y_j^T B^T - B Y_j < 0 \quad (16)$$

for $h_i(\theta(t)) h_j(\theta(t)) \neq 0, \forall t \geq 0$, and $i, j = 1, 2, \dots, s$.

The set of control law gain matrices is given as follows

$$K_j = Y_j X^{-1}, \quad j = 1, 2, \dots, s \quad (17)$$

IV. ILLUSTRATIVE EXAMPLES

Example I.

Considering the system model (9), (10) parameter matrices A, B, C and C_z are as follows

$$A_i = \begin{bmatrix} 0 & 1 & 0 & 0.05 & 0 & 0.05c_k \\ 2 & -1 & 0.08 & 0.01 & 0.08c_k & 0.01 \\ 0 & 0.05 & 0 & -1 & 0 & 0.05 \\ 0.08 & 0.01 & -2 & 0.5 & 0.08 & 0.01 \\ 0 & 0.05 & 0 & 0.05 & 0 & 1 \\ 0.08 & 0.01b_l & 0.08 & 0.01 & 1 & -2 \end{bmatrix},$$

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad C = I_{6 \times 6}, \quad C_z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

with the associations

$$\begin{aligned} i = 1 &\leftarrow (l = 1, k = 1) & i = 2 &\leftarrow (l = 2, k = 1) \\ i = 3 &\leftarrow (l = 1, k = 2) & i = 4 &\leftarrow (l = 2, k = 2) \end{aligned}$$

The variables $p(t) \in \langle 0, 3 \rangle$ and $q_3(t) \in \langle -1, 4 \rangle$ are bounded on the prescribed sectors then vector of the premise variables can be chosen as follows

$$\theta(t) = \begin{bmatrix} \theta_1(t) & \theta_2(t) \end{bmatrix} = \begin{bmatrix} q_3(t) & p(t) \end{bmatrix}$$

Thus, the set of nonlinear sector functions

$$w_{11}(q_3(t)) = \frac{b_1 - q_3(t)}{b_1 - b_2}, \quad b_1 = -1, \quad b_2 = 4$$

$$w_{12}(q_3(t)) = \frac{q_3(t) - b_2}{b_1 - b_2} = 1 - w_{11}(q_3(t))$$

$$w_{21}(p(t)) = \frac{c_1 - p(t)}{c_1 - c_2}, \quad c_1 = 0, \quad c_2 = 3$$

$$w_{22}(p(t)) = \frac{p(t) - c_2}{c_1 - c_2} = 1 - w_{21}(p(t))$$

implies the next set of normalized membership functions

$$h_1(q_3(t), p(t)) = w_{11}(q_3(t))w_{21}(p(t))$$

$$h_2(q_3(t), p(t)) = w_{12}(q_3(t))w_{21}(p(t))$$

$$h_3(q_3(t), p(t)) = w_{11}(q_3(t))w_{22}(p(t))$$

$$h_4(q_3(t), p(t)) = w_{12}(q_3(t))w_{22}(p(t))$$

The values of parameters b and c are the maximal values of the signals $q_3(t)$ and $p(t)$ experimental obtained. Signal $p(t)$ was modeled as in Fig. 1.

LQ controller design

The controller was designed for linear dynamic system and applied to nonlinear system.

Using the variables $b_l = 1.5, c_k = 1.5$ and Matlab function $[K, P, E] = \text{lqr}(A, B, Q, R)$ to LQR method design there was computed the controller gain matrix K

$$K = \begin{bmatrix} 4.235 & 2.240 & 0.195 & 0.012 & 0.145 & 0.106 \\ 0.194 & 0.012 & -4.235 & 3.623 & 0.162 & -0.003 \\ 0.145 & 0.106 & 0.153 & -0.003 & 2.415 & 1.139 \end{bmatrix}$$

Note, $Q^{n \times n}$ and $R^{m \times m}$ are identity matrices.

The simulation results are shown in Fig. 2.

Example II.

Matrices B, C, C_z, Q, R and membership functions were used same as in example I. and matrix A is as follow

$$A = \begin{bmatrix} 0 & 1 & 0 & 0.05 & 0 & 0.05 \\ 2 & -1b_l & 0.08 & 0.01 & 0.08 & 0.01 \\ 0 & 0.05 & 0 & -1c_k & 0 & 0.05 \\ 0.08 & 0.01 & -2 & 0.5 & 0.08 & 0.01 \\ 0 & 0.05 & 0 & 0.05 & 0 & 1 \\ 0.08 & 0.01 & 0.08 & 0.01 & 1c_k & -2 \end{bmatrix}$$

A. LQ controller design

For the controller gain design were used the variables $b_l = 0.2, c_k = 0.2$ and using Matlab function $[K, P, E] = \text{lqr}(A, B, Q, R)$ there was computed the matrix K

$$K = \begin{bmatrix} 4.229 & 2.887 & 0.280 & 0.061 & 0.129 & 0.101 \\ 0.281 & 0.061 & -4.226 & 2.228 & 0.211 & 0.014 \\ 0.126 & 0.101 & 0.221 & 0.014 & 1.217 & 0.731 \end{bmatrix}$$

The simulation results are shown in Fig. 3.

The controller designed for this linear dynamic system and applied to nonlinear system is insufficient as is evident from Fig. 3. It is necessary to design the fuzzy state control for this nonlinear system.

B. Fuzzy control design

Thus, solving (15), (16) with respect to the LMI matrix variables X and $Y_j, j = 1, 2, 3, 4$ using Self-Dual-Minimization (SeDuMi) package for Matlab and according to (17) matrices K_1, K_2, K_3, K_4 were obtained

$$K_1 = \begin{bmatrix} 7.093 & 15.609 & 0.421 & -0.068 & 0.101 & 0.061 \\ 0.216 & -0.131 & -5.342 & 9.275 & 0.735 & 0.040 \\ -0.077 & 0.192 & 0.319 & 0.897 & 27.764 & 13.995 \end{bmatrix}$$

$$K_2 = \begin{bmatrix} 7.017 & 15.340 & 0.413 & -0.058 & 0.166 & 0.094 \\ 0.196 & -0.158 & -5.314 & 9.202 & 0.772 & 0.059 \\ -0.073 & 0.184 & 0.316 & 0.867 & 27.084 & 13.579 \end{bmatrix}$$

$$K_3 = \begin{bmatrix} 7.108 & 15.642 & 0.435 & -0.121 & -0.203 & -0.121 \\ 0.183 & -0.221 & -5.473 & 9.622 & 0.919 & 0.134 \\ -0.113 & 0.057 & 0.300 & 0.933 & 27.474 & 13.804 \end{bmatrix}$$

$$K_4 = \begin{bmatrix} 6.947 & 15.107 & 0.413 & -0.071 & 0.137 & 0.065 \\ 0.211 & -0.153 & -5.415 & 9.471 & 0.859 & 0.114 \\ -0.109 & 0.083 & 0.281 & 0.992 & 27.807 & 14.003 \end{bmatrix}$$

and were applied to nonlinear Takagi-Sugeno-Kang fuzzy model (9), (10).

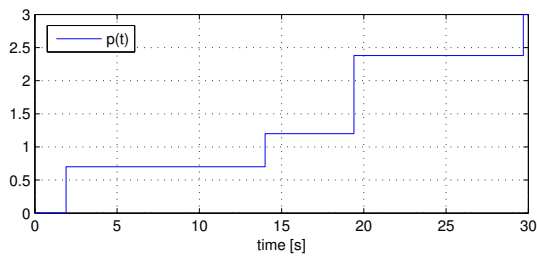
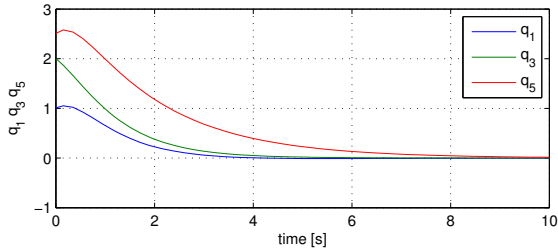
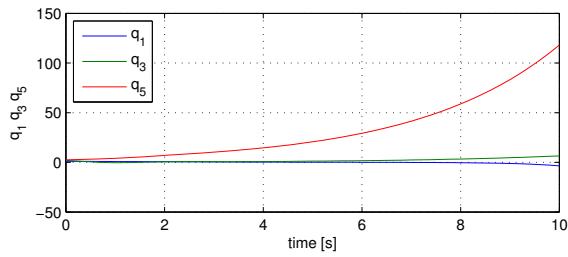
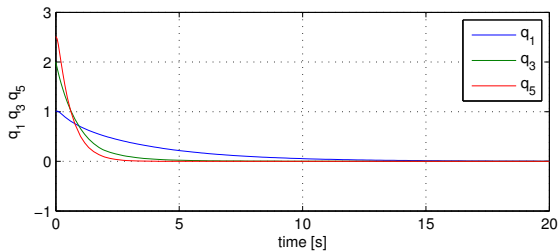
Simulation results (Fig. 4.) shows on the effectiveness the fuzzy state controller application for the nonlinear systems.

V. CONCLUSION

The paper presents the stabilizing controller application for a class of continuous-time nonlinear systems represented by Takagi-Sugeno-Kang models, and measurable premise variables. Simulation results illustrate the design procedure and shows on the effectiveness the fuzzy state controller application for the nonlinear systems.

ACKNOWLEDGMENT

The work presented in this paper was supported by VEGA, Grant Agency of Ministry of Education and Academy of Science of Slovak Republic under Grant No. 1/0256/11, as well as by Research & Development Operational Programme Grant No. 26220120030 realized in Development of Center of Information and Communication Technologies for Knowledge Systems. These supports are very gratefully acknowledged.


 Fig. 1. Fault $p(t)$

 Fig. 2. Control of nonlinear system from *example I* using linear controller with initial conditions $q_1 = 1, q_3 = 2, q_5 = 2.5$

 Fig. 3. Control of nonlinear system from *example II. a)* using linear controller with initial conditions $q_1 = 1, q_3 = 2, q_5 = 2.5$

 Fig. 4. Control of nonlinear Takagi-Sugeno-Kang system from *example II. b)* using fuzzy controller with initial conditions $q_1 = 1, q_3 = 2, q_5 = 2.5$

REFERENCES

- [1] I. Abdelmalek, N. Golea, and M.L. Hadjili, "A new fuzzy Lyapunov approach to non-quadratic stabilization of Takagi-Sugeno-Kang fuzzy models", *International Journal of Applied Mathematics and Computer Science*, vol. 17, 2007, no. 1, pp. 3951.
- [2] B. Boyd, L. El Ghaoui, E. Peron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, SIAM Society for Industrial and Applied Mathematics, 1994.
- [3] M. Johansson, A. Rantzer, and K.E. Arzen, "Piecewise quadratic stability of fuzzy systems. *IEEE Transactions on Fuzzy Systems*, vol. 7, 1999, pp. 713722.
- [4] F. Khaber, K. Zehar, and A. Hamzaoui, "State feedback controller design via Takagi-Sugeno fuzzy model: LMI approach", *International Journal of Computational Intelligence*, vol. 2, 2005, no. 3, pp. 148-153.

- [5] D. Krokavec and A. Filasová, *Dynamic Systems Diagnosis*, Košice, Elfa, 2007. (in Slovak)
- [6] D. Krokavec and A. Filasová, "Equality constraints in sensor faults re-configurable control design", *Proceedings of the 8th European Workshop on Advanced Control and Diagnosis ACD 2010*, Ferrara, Italy, 2010, pp. 184-189.
- [7] D. Krokavec, A. Filasová and V. Hladký, "Stabilizing fuzzy control for a class of nonlinear systems", *Proceedings of the 10th IEEE International Symposium on Applied Machine Intelligence and Informatics SAMI 2012*, Herlany, Slovakia, 2012, 53-58.
- [8] K. Michels, F. Klawonn, R. Kruse, and A. Nürnberger, *Fuzzy Control. Fundamentals, Stability and Design of Fuzzy Controllers*, Berlin, Springer-Verlag, 2006.
- [9] K.M. Passino and S. Yurkovich, *Fuzzy Control*, Berkeley, Addison-Wesley Longman, 1998.
- [10] R.E. Precup, S. Preitl, E.M. Petriu, J.K. Tar, M.L. Tomescu, and C. Pozna, "Generic two-degree-of-freedom linear and fuzzy controllers for integral processes", *Journal of The Franklin Institute. Engineering and Applied Mathematics*, vol. 346, 2009, no. 10, pp. 988-1003.
- [11] R.E. Precup, S. Preitl, I.J. Rudas, M.L. Tomescu, and J.K. Tar, "Design and experiments for a class of fuzzy controlled servo systems", *IEEE/ASME Transactions on Mechatronics*, vol. 13, 2008, no. 1, pp. 2235.
- [12] T. Takagi and M. Sugeno, "Fuzzy Identification of Systems and Its Applications to Modeling and Control", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, 1985, no. 1, pp. 116132.
- [13] K. Tanaka and H.O. Wang, *Fuzzy Control Systems Design and Analysis. A Linear Matrix Inequality Approach*, New York, John Wiley & Sons, 2001.
- [14] H.O. Wang, K. Tanaka and M.F. Griffin, "An approach to fuzzy control of nonlinear systems: Stability and design issues", *IEEE Transactions on Fuzzy Systems*, vol. 4, 1996, no. 1, pp. 14-23.
- [15] F. Yang and Y. Li, "Set-membership fuzzy filtering for nonlinear discrete-time systems", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, 2010, no. 1, pp. 116-124.
- [16] F. Yang and Y. Li, "Set-membership fuzzy filtering for nonlinear discrete-time systems", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, 2010, no. 1, pp. 116-124.

Gesture recognition using Static Bayesian Tree algorithm

¹Peter ADAMONDY (2nd year of master degree), ²Peter PAPCUN (1st year)

Supervisor: ³Ján Jadlovský

^{1,2,3}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹peter.adamondy@student.tuke.sk, ²peter.papcun@tuke.sk, ³jan.jadlovsky@tuke.sk

Abstract – We propose method to recognize movement gestures based on Static Bayesian Classification Tree algorithm. This algorithm is merge of two classification methods. It inherits naive Bayesian classifier and implements it into predefined decision tree structure. Every node of this tree is independent Bayesian classifier trained only to distinguish between lower level nodes. This algorithm is very fast to compute and thanks to Bayesian probabilistic model it is also very robust.

Keywords — Bayesian classification, decision tree, gesture recognition.

I. INTRODUCTION

Human-machine interaction designs, as described by Guy A. [1], are in the present times shifting from the machine centered design (MCD) to the human centered design (HCD). The main idea of HCD is to take control of a machine as close as possible to human nature. Control of computer based systems relies strongly on the cognitive skills of their users. They need to learn how to control software and hardware to finish given task. HCD allows them to decrease learning time to minimum. Correct recognition of nature human behavior and response in adequate manner is great part of HCD. Here we will focus on dynamic human gesture recognition.

II. OVERVIEW

Our aim was to create simple, fast and robust algorithm that can be used not only for gesture recognition, but for various classification problems as well. For purposes of this paper we will focus on gesture recognition only.

First we need to specify what we mean by “gesture”. Gesture is deterministic movement of human body part or parts. For example, if you want to say “hello” to someone, you can express that through waving gesture of your hand. Human brain recognizes that gesture and can respond to it accordingly i.e. wave back. In computer systems we use one or multiple cameras to catch visual objects and points of interest (POI). In this case POI is human hand. For standard cameras and computers is extracting of hand position (skeletal point) time demanding. We decided to use Kinect sensor that already has this ability. Computer will be employed only by gesture recognition task.

Static Naive Bayesian Tree algorithm (SNBT) or shorter Static Bayesian Tree (SBT) consists of three parts and its implementation can be divided to two stages. First part of

algorithm is data extraction in form of vectors. After extracting of x, y and z coordinates, those vectors are passed to preprocessing stage of algorithm. We need to extract several features that will determine characteristics of specified gesture. During preprocessing stage we will merge all dimensions into one vector. From merged vector several features will be extracted:

- Mean crossing rate (MCR)
- Short time average energy (STAE)
- Standard deviation
- Mean

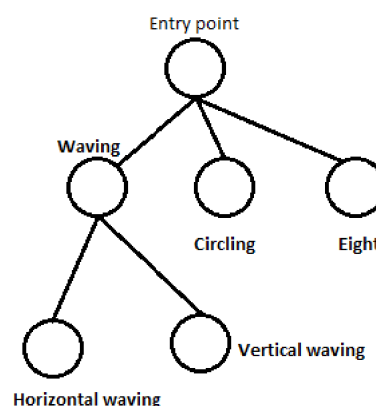


Fig. 1: Decision tree structure with Bayesian classifier nodes.

Those features are enough to correctly classify specific gestures. For our example four different gestures have been used:

- Horizontal waving
- Vertical waving
- “Eight”
- Circling

Those classes are structured into decision tree as seen on Fig. 1. All classes with their features are passed into learning stage of algorithm where mean and variance of features for each node is computed, so that they can be used later in classification stage.

Classification part of algorithm is of course separated and it will use precomputed parameters from the learning stage. Classified sample sets must be also preprocessed. Here we are

extracting exactly same features as in preprocessing stage of learning part of algorithm.

III. PREPROCESSING AND FEATURE EXTRACTION

As mentioned above getting x , y and z coordinate vectors of POI must be preprocessed and few features must be extracted from them. Firstly vectors are merged into one by following formula:

$$f_i = \sqrt{(x_i^2 + y_i^2 + z_i^2)} \quad (1),$$

where f_i is merged value vector, and x_i , y_i , z_i are coordinate vectors. Significant amount of noise is always present when collecting data from sensors. Kinect is no exception. To cut off high frequency noise from merged vector, Butterworth low pass filter is applied [3].

After this procedure we can now extract features mentioned above. Mean crossing rate (MCR) is calculated as rate of crossing over mean value of given vector. Frequency is in direct proportion to MCR. Second feature is short time average energy (STAE). STAE is calculated by formula (2).

$$s_i = \sum |(\mathcal{F}(f_i^2))| \quad (2),$$

where, s_i is STAE value, f_i^2 is square values of merged vector and \mathcal{F} is Fourier transform of given vector. Mean μ_c as third value if self explanatory, as it is mean of all values in given vector. Last feature is standard deviation calculated by formula (3).

$$s_{dev} = \sqrt{\frac{1}{N} - 1 \sum_{i=1}^N (x_i - \bar{x})^2} \quad (3),$$

where s_{dev} is standard deviation, N is count of samples in given vector f_i , x_i is i -th sample and \bar{x} is mean value.

IV. BACKGROUND AND CALCULATIONS

Let $F = (e_1, \dots, e_n)$ be a vector of features, where each feature takes values from its domain D_i as described in [2]. The set of all feature vectors is denoted $\Omega = D_1 \times \dots \times D_n$. Let C be an unobserved random variable denoting *class* of an example, where C can take one of m values $c \in \{0, \dots, m-1\}$.

A function $g: \Omega \rightarrow \{0, \dots, m-1\}$, where $g(x) = C$, denotes a *concept* to be learned. Deterministic $g(x)$ corresponds to a concept without noise, which always assigns the same class to a concept example (e.g., disjunctive and conjunctive concept are deterministic).

A *classifier* is defined by a (deterministic) function $h: \Omega \rightarrow \{0, \dots, m-1\}$ (a *hypothesis*) that assigns a class to any given example. Common approach is to associate each class i with a discriminant function $f_i(x)$, $i = 0, \dots, m-1$, and let the classifier select the class with maximum discriminant function on a given example $h(x) = \arg \max_i f_i(x)$. The Bayes classifier uses as discriminant functions the class posterior probabilities given feature vector F . Applying Bayes rule gives

$$P(C=i|X=F) = \frac{P(X=F|C=i)P(C=i)}{P(X=F)} \quad (4),$$

where $P(X=F)$ is identical for all classes, and therefore can be ignored. This yields Bayes discriminant functions

$$f_i^x(F) = P(X=F|C=i)P(C=i) \quad (5),$$

where $P(X=F|C=i)$ is called as class-conditional probability distribution (CPD). Thus, the Bayes classifier finds the maximum *a posteriori* probability (MAP) hypothesis given example F .

Direct estimation of $P(X=F|C=i)$ from a given set of training examples is hard when the feature space is high-dimensional. Therefore using simplifying assumption that features are independent given the class. Another simplifying assumption is made as we assume that *a priori* probability of certain class is same for all classes. In other words we assume *equiprobability* of classes. We have no reason for this assumption so it might be bad idea. In order to decrease error risk distributed through learning we use only binary tree structure.

We are dealing here with continuous data sets. Typical assumption for this type of data is that values associated with each class are distributed according to Gaussian distribution. We firstly segment data by class, and then compute mean and variance of feature e_i in each class. Since we use tree structure, every analyzed vector e_i contains data from nodes or leaves of subtree that lies under currently calculated node. Probability of some value given a class $P(e_i=v/c)$ can be then computed by formula (6).

$$P(e_i=v|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}} \quad (6),$$

where $P(e_i=v/c)$ is probability of some value given class, μ_c is mean of the values in e_i associated with class c and σ_c^2 is variance of the values in e_i in given class c .

Class with highest probability is considered as given class and therefor is assumed as currently performed gesture.

V. GESTURES

As mentioned in section II. of this paper we are using four gestures. In this section we will focus on explaining each of them.

Horizontal waving movement is shown on Fig. 2. It is cyclic movement of POI between P1 and P2.



Fig. 2: Horizontal waving of POI

Vertical waving movement is shown on Fig.3. It is similar to the horizontal waving.



Fig. 3: Vertical wave of POI

Previous gesture types were pretty self explanatory. But following gestures need a little explanation. On Fig. 4 is circling movement of POI. We do not care about starting point of circling since we have implemented value vector shifting. Each collected frame of data are overlapping previous frame with specified amount of samples. With feature extraction from merged frames and strongly specifiable STAE parameter that is extracted from Fourier transform of data frames, we no longer need to care where will gesture start or end. This fact gives us high robustness for classification.

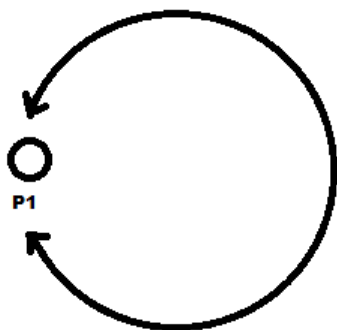


Fig. 4: Circling movement of POI.

The last gesture is so called “eight” which is similar to circular movement. Its scheme is drawn in Fig. 5.

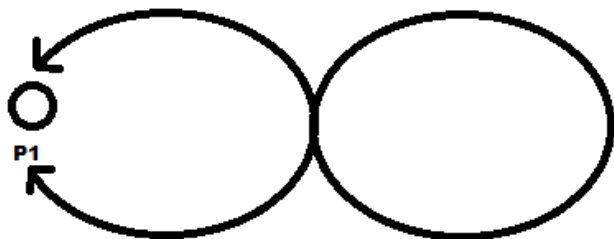


Fig. 5: “Eight” circling movement of POI.

We use only four types of gestures, but any number of gestures can be added as required. Bayesian classifier has several properties that make it very useful. In particular, the decoupling of the class condition feature distribution means that each distribution can be independently estimated as a one

dimensional distribution. Combining with static tree structure it will eliminate problem with exponentially decreasing accuracy with increasing count of classes. Also more feature vectors can be added, that can increase accuracy of algorithm.

VI. CLASSIFICATION

To determine effectiveness of this classification algorithm we will use confusion table. In each cell of this table will be amount of data frames classified as probable class. We used 1000 samples for each class. If sample is classified properly it is added to cell where column and row has same name. Otherwise classification failed.

	horizontal wave	vertical wave	circling	„eight“	accuracy
horizontal wave	940	60	0	0	94,0000%
vertical wave	79	893	28	0	89,3000%
circling	0	42	872	86	87,2000%
„eight“	0	2	72	928	92,8000%
overall accuracy					90,8250%

Tab. 1: Confusion matrix of SNBT algorithm.

VII. CONCLUSION AND FUTURE WORK

Static Naive Bayesian Tree algorithm is easy fast and robust. As we shown in this paper it is also accurate with overall accuracy around 90.825% of properly classified samples. We found that significant impact on accuracy has size of data frame and size of overlapping vector. SNBT can be used for classification problems where classes are structured into static tree. Further research time should be focused on implementing Bayesian classifiers into dynamic tree structure which will be created during learning stage of algorithm.

In future, gesture recognition will be used for starting specific procedure. For example we are using kinect to control Mitsubishi robot. Robot is copying movement of human hand. We can use other hand to perform specific gestures, that will for example start recording of the robot movement or start some device connected on the robot's universal device slot (i.e. welding machine).

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

This work has been supported by the Scientific Grant Agency of Slovak Republic under project Vega No.1/0286/11 Dynamic Hybrid Architectures of the Multiagent Network Control Systems.

REFERENCES

- [1] Boy, G. A., “Perceived Complexity and Cognitive stability in Human-Centered design”, Proceedings of the HCI International 2007 Conference, Beijing, China., (2007).
- [2] Rish, I., “Empirical study of the naive Bayes classifier”. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, (2001).
- [3] Butterworth S., “On the Theory of Filter Amplifiers”, In Wireless Engineer, vol.7, pp. 536-541, (1930)

GPGPU: The Increase of the Computing Power for Data Encryption

¹Michal ENNERT (1st year), ²Peter Ivančák (1st year)
 Supervisor: ³Liberios Vokorokos

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹michal.ennert@tuke.sk, ²peter.ivančák@tuke.sk, ³liberios.vokorokos@tuke.sk

Abstract— In recent times, GPUs have been considered a potential source of powerful computational power for non-graphical applications. This fact can be used for data encryption. This paper creates an overview of the most widely used technologies based on GPGPU. It presents current overview of use of GPGPU for data encryption and performance maximization in GPU calculations.

Keywords—GPGPU, GPU, encryption, OpenCL, CUDA.

I. INTRODUCTION

The idea to use graphics hardware as a math coprocessor to the main central processing unit (CPU) has become the object of intense research at the end of 2002. At that time, was first introduced and used acronym GPGPU - General Purpose Computing on Graphics Processors. Thus began the era of practical implementation of the computing power of graphic processors (GPU) in specific applications. [1].

At present, the CPU and GPU architecture complement each other. CPU takes care of algorithm and GPU provides data processing. GPUs are generally very effective especially in parallel data operations. The computing power of GPU is in this case several times greater than the CPUs. The main reason of such performance is specialization of GPU processor to processor intensive, highly parallel computing. The core of GPUs is parallel SIMD processor type architecture, which computing power now accounts for up to several TFLOPS. [2].

More specifically, the GPU processors are suitable for efficient solution of problems which can be expressed as data-parallel computing. This means that the same task is performed parallelly on multiple data elements SIMD, where the number of arithmetic operations is greater than the number of memory operations. Because the same program is performed for each data element, the need for sophisticated management of the flow of operations is lower. The latency of memory access through a large cache registers can be compensated just by calculations with high arithmetic intensity [3].

It is advisable to use this computing thesis for a different encryption methods in which parallelism can be applied. The use of GPUs to speed up the computation of cryptographic primitives was pioneered by D. Cook al. in [4].

Consequently, it makes sense to employ a GPU as an accelerator for cryptography given its computing power. The goal of this paper is to create an overview of the most widely used technologies based on GPGPU. Another task is to present a way to maximize performance in GPU calculations. The final task is to provide a current overview of GPGPU applications for encryption techniques in preparation for further research in the potential of GPU in cryptography.

II. ANALYSIS OF CURENT SITUATION OF GPGPU

A. GPGPU architecture

GPU became much more complex than CPU processors, because of concentrating just on stream processing based on SIMD architectures. In terms of GPGPU applications to graphics hardware, it is composed of GPU and video memory and it appears to be a math coprocessor to the main processor CPU. CPU is in charge of running the main application and performs the operations as part of the reservation system memory for the application, create GUI applications, etc. Using the control respectively control commands located in libraries or video hardware drivers can control the correct functioning of the GPU processor for performing data - parallel operations [5].

GPU-based architecture, because of implementing large number of parallel units in conjunction with high-speed memories with high throughput, is a base not only for graphic operations, but can be also found in other implementations. GPU architecture compared to CPU architecture favors more transistors designed for data processing (ALU), than for their caching (Cache) or to check their current implementation (Control) [2] as is in Fig. 1.

B. GPU programming and programming languages

Today's GPUs are general-purpose parallel processors with support for accessible programming interfaces and industry-standard languages. Developers who port their applications to GPUs often achieve speed ups of orders of magnitude vs. optimized CPU implementations. Parallel programming can improve program performance. Not all codes are good candidates for parallel computation. [6].

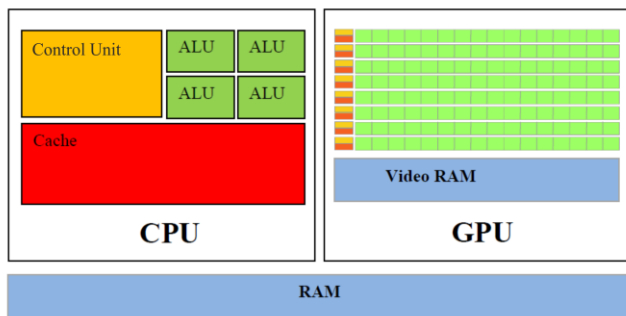


Fig. 1. Comparison between CPU and GPU architecture [5].

Nowadays even complex GPGPU applications are made. A group of languages directly made by producers of GPUs are being made, for programming graphics hardware GPU processors as processors to carry out general calculations. Applications written in standard programming languages (C, C++ and others) which want to use all possibilities of GPU, must be able to communicate with graphics drivers. This is enabled by API, which eliminates low-level programming of a display device. It makes programming much easier and much more affordable. It also allows using of all possibilities of GPU. Currently the most widely used APIs are Direct3D and OpenGL. Both standards have the same conception of graphic process. They are historically the oldest and also the lowest platforms in this area of programming. For programming of the vertex / fragment / geometry computing units were subsequently developed separate shader programming languages which fully cooperate with different graphics APIs. For Direct3D a language HLSL (High Level Shader Language) has been made and with OpenGL cooperates language GLSL (OpenGL Shading Language). Shader language compatible with both graphical API was developed by nVidia and it is called Cg (C for Graphics). Paradoxically, the strong restriction of multiplatform programmability of GPU processors and generally the graphic hardware as well is just the existence of two parallel graphics standards OpenGL and Direct3D. Even portability of code between multiple platforms is possible only through a graphical OpenGL, but that is not flexible enough to implement new technologies into the use. GPGPU applications use meta-languages with syntax similar to C or C++. The meta-languages are in many cases a sort of extension to the graphical standards and in many cases also supporting bases for programming applications based on GPGPU [5][7].

C. GPGPU languages comparison

The main task is to facilitate GPGPU programming language with syntax conversion of C or C++ code into shader language or code of some of the graphic standards. This transformation is performed by the CPU processor. The two leading companies in the development of graphics cards, NVIDIA and ATI offer their own solutions.

nVIDIA CUDA

The older and most widely used is the meta-language for programming GPU called CUDA made by NVIDIA. CUDA (Compute Unified Device Architecture) is a proprietary option for programming GPU respectively unified shader units using a set of features and support libraries. CUDA is a C++

compiler, which also contains several language extension of C++ that can work directly GPU. These extensions are supported by software libraries and special CUDA driver. The disadvantages include CUDA poor support for the management of difficult calculations for multiple graphics adapters, connected parallel in computer system using SLI technology. The programming model is based on a hierarchy of groups of threads, shared memory and synchronizing threads. It provides several types of parallelism, but explicitly requires distributing problem into sub-problems that can be solved independently in parallel. This decomposition provides scalability of CUDA architecture into the graphics cards, which differs in numbers of processors and memory capacity. CUDA API provides two solutions for the creation of programs: high-level CUDA Runtime API and low-level CUDA driver API. Their mutual use is eliminated. CUDA Driver API is more complex and harder to use, but provides more flexibility. Both API work with Direct3D and OpenGL. Language CUDA is an extension of C language, designed for writing code for GPU. The CUDA architecture uses vocabulary specific to the technology. It should be noted some of the concepts that will be mentioned later:

- thread – basic data element,
- warp – group of 32 threads, which is the smallest amount of data in SIMD architecture processed by CUDA multiprocessor,
- block – can contain 64 - 512 threads, it is used for efficiency,
- grid – group of blocks, which are processed at once,
- host - CPU machine,
- device – GPU device [5].

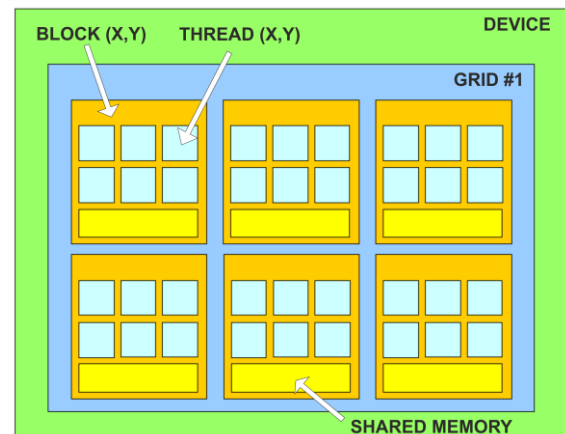


Fig. 2. CUDA kernel architecture organization [9].

The application consists of a main program that is performed on the host and CPU and CUDA program that is performed on GPU accelerator. The host program calls computational core (kernel), or more kernels, which is then performed on the GPU. Kernel operations are performed in parallel with the fibers in the GPU. Optimizing CUDA program consists of finding the optimal balance between the number of blocks and their sizes.

AMD/ATI STREAM

The competitor technology for programming GPU from AMD / Ati is called STREAM. In essence, it is similar to CUDA language. It is built on the historic old CTM system, which was a proprietary GPU programming option with direct

access to GPU instructions given by the assembler. One of its advantages is the support of management calculations for multiple graphics adapters, connected in parallel in a computer system with CrossFire technology. ATI Stream is essentially a set of advanced technical (graphics processor) and software (software interface) solutions, which allow using and synchronizing the CPU and GPU.

OpenCL

Open Computing Language (OpenCL) is a cross-platform technology standard in development under the Khronos Group. It represents the first attempt to unify application development environment for GPGPU although in essence it is designed as a multiplatform heterogeneous solution for writing applications using parallel computing process. OpenCL provides a uniform programming environment for developers of software. OpenCL is an open industry standard developed by Apple Inc. and managed by Khronos Group. It is supported by main hardware manufacturers, including ATI and nVidia. Standard OpenCL consists of APIs designed for management of parallel computing and programming language that is specifically designed to enter these calculations. OpenCL is an open standard for general purposes of parallel programming of CPUs, GPUs and other processors, which brings an efficient approach to perform these heterogeneous processor platforms. By creating an effective, low-level programming interface, OpenCL will form the base layer for parallel computing ecosystem of platform-independent tools, middleware and applications.

OpenCL consists of a host application and the OpenCL program to be carried out on computing device. Host program activates computing kernels contained in the device program. Kernels are executed in parallel on multiple device data units through the process elements. OpenCL kernels are written in OpenCL C programming language. Host program controls device using OpenCL API. Graphical operations are provided by OpenGL. As OpenCL not focus only on the GPU, but also in multinucleated CPU OpenCL kernel distinguishes between two types:

- data-parallel, which are suitable for GPU architecture,
- task-parallel, which are suitable for CPU architecture [8].

CUDA vs. OpenCL

For almost all types of calculations, all accelerators in practice showed that CUDA can effectively use the potential of computing accelerators NVIDIA. So is confirmed the presumption that OpenCL will lose because of its universality and vice versa, CUDA, along with hardware, because it is better customized and makes better use of its potential. CUDA is also easier to learn and code, way more common, enjoys a huge range of available libraries for dedicated domains. OpenCL, while having a steeper learning curve and a reputation of being more involved to code for, is an open standard which is likely to gain more and more acceptance over time. OpenCL is a layer on top of the CUDA driver interface. As such, OpenCL is one avenue to GPU computing through CUDA, CUDA C is another (see Fig. 3.).

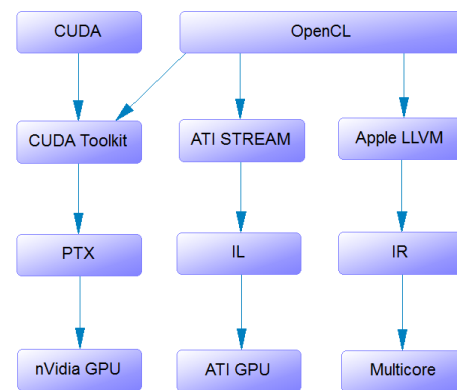


Fig. 3. CUDA vs. OpenCL

CUDA PROS:

- C-like language called C for CUDA for writing the kernel code. Kernel code has full pointer support
- Supports C++ constructs such as templating.
- Fairly simple integration API.
- Better fully GPU accelerated libraries are currently available
- Larger assortment of higher quality bindings for various languages, both commercial and free.
- Very well documented, amount of samples available for various platforms.
- Many debugging and visual profiling tools.
- Updates are more regular.

CUDA CONS:

- Restricted to Nvidia GPUs only
- Will not fall back on to CPU if CUDA accelerated hardware is unavailable.

OPENCL PROS:

- Much wider range of hardware and platform support. Supports AMD, nVidia and Intel GPU sequally. Can also be used on newer versions of android phones, iPhones and other devices.
- Can fall-back on to CPU if GPU support does not exist.
- Supports synchronization over multiple devices.
- Easy to get started with integrating OpenCL kernels into code.
- An open standard and not vendor locked.
- Kernell based on C99 specification
- Share resources with OpenGL
- Intel and IBM expected soon to mature their drivers

OPENCL CONS:

- Lacks mature libraries
- Debugging and profiling tools are not as advanced as CUDAs.

III. GPGPU AND DATA ENCRYPTION

Currently, it is already used GPGPU technology in several areas. As mentioned earlier, GPU processors are only effective in solving problems that can be processed as stream, in parallel. From this fact result the possibility of using GPGPU in cryptography (encryption and decryption of data). Data encryption has always been a key point in warranting confidentiality, but encrypting large data imposes a significant computation load on the CPUs. A way to reduce this computational load is to employ GPGPU technology [10].

GPU technology can be used effectively to improve the

computational performance of cryptographic algorithms. Effective use mainly provide high-parallelized algorithms like the popular symmetric encryption AES, DES but also others. If a sufficiently large amount of data is used, algorithms can significantly reduce the time required for encryption [11].

Thus, there is potential to use the processing power GPU technology similar but more effective way than the previous CPU. The main requirements for the calculation of high performance include high throughput. It was demonstrated that the GPU performs best using large packet sizes and thus suits applications which require bulk data encryption/decryption [12].

IV. GPU COMPUTING POWER MAXIMIZATION FOR ENCRYPTION

The problem in the calculations can be CPU load. The primary task for the CPU is the data transfer and initialization GPU. The CPU occupancy has no influence on data processing in the GPU. However, total time can be adversely affected. The performance improvements achieved using GPGPU technology are dependent on some general principles that should be respected. The overall goal is to maximize occupancy of the GPU such that each multiprocessor has the maximum possible number of threads queue data anytime. It is important to use memory efficiently, while minimizing the amount of branching on the threads. Unlike the CPU, which plans to and calculates changes in fiber branching, GPUs can do some branching only after some amount of calculations. This can cause some threads in the calculation will not be used. Finally, it is important to reduce the number and size of data transfers between CPU and GPU [13]. From the experience gathered so far on research using GPU encoding method results a number of practical recommendations for programming algorithms:

Maximize use of available processing power:

- Maximize independent parallelism in the algorithm to enable easy partitioning in threads and blocks.
- Keep resource usage low to allow concurrent execution of as many threads as possible, i.e., use only a small number of registers per thread and shared memory per block.
- Maximize arithmetic intensity, i.e., match the arithmetic to bandwidth ratio to the GPU design philosophy: GPUs spend their transistors on ALUs, not caches. Bearing this in mind allows hiding memory Access latency by the use of independent computations (latency hiding).
- Avoid divergent threads in the same warp.

Maximize use of available memory bandwidth:

- Avoid memory transfers between host and device by shifting more computations from the host to the GPU.
- Use shared memory instead of global memory for variables.
- Use constant or texture memory instead of global memory for constants.
- Coalesce global memory accesses, i.e., choose Access patterns that allow to combine several accesses in the same warp to one, wider access.
- Avoid bank conflicts when utilizing shared memory, i.e., choose patterns that result in the Access of different banks per warp.
- Match Access patterns for constant and texture memory to

the cache design [14].

V. CONCLUSION

Modern approach in the world of computing as well as the ever-escalating trend is the deployment and subsequent use of a parallel environment for all calculations, which show the possibility of parallel processing. It's just the two factors. The first important factor is that the concept of sequential implementation of operations is no longer just simply technologically outdated, but very powerful and deficient in specific cases literally been unsatisfactory. The second, also non-negligible factor is that technological advances over the past few years has made available the possibility of using high performance and parallel implementation of operations by professional public community. Because of the nature of different levels of programming, job demands and arithmetic - logical, numerical and graphical operations, it is necessary to consider what form of parallelization will be applied at what time of data handling.

The major obstacle lies in the data transfer between the graphics card and main memory. If the GPU calculation deals with a small content of the data, it is more effective to use CPU. Implementation of encryption algorithms on the GPU is not an easy task. Optimization algorithms in GPU environment remains a great challenge.

GPGPU technology is a powerful tool that offers the ability of parallel massive data processing. However, only some highly parallel encryption algorithms can effectively exploit its potential parallelism. Today, there are few results, and comparisons of the performance of various encryption methods. They show a trend to reduce computation time by limiting the data transfer between CPU and GPU, which can have a negative effect on GPGPU technology performance alone. A good solution can be made by building a universal criteria and parameters for optimal computational model multiple encryption methods.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0008-10.

REFERENCES

- [1] Pharr M., Fernando R.: GPU Gems 2: Programming Techniques for High-Performance Graphics and General-Purpose Computation, Addison-Wesley Professional, 2005, ISBN 0321335597.
- [2] NVIDIA Corporation: NVIDIA CUDA Compute Unified Device Architecture – Programming Guide [online], 2010, Available: <http://developer.nvidia.com/cuda>.
- [3] Sobota B., Perhák J., Hrozek F.: Využitie technológií GPGPU pri vizualizáciách, simuláciách a používateľských rozhraniach, 2010, Technická univerzita v Košiciach.
- [4] Debra L. Cook, John Ioannidis, Angelos D. Keromytis, and Jake Luck. Crypto Graphics: Secret Key Cryptography Using Graphics Cards. In Alfred Menezes, editor, CT-RSA, volume 3376 of Lecture Notes in Computer Science. Springer, 2005.
- [5] Sobota B., Korečko Š., Látka L., Szabó C., Hrozek F.: Riešenie úloh spracovania rozsiahlych grafických údajov v prostredí paralelných počítačových systémov, 2012, Edičné stredisko TU, Košice, ISBN 978-80-553-0864-7.
- [6] Harris Mark: GPGPU [online], 2012, Available: <http://gpgpu.org>.
- [7] Vokorokos Liberios, Perhák Ján, Madoš Branislav: Distributed GPGPU – Implementation Possibilities, Proc. of 6th International Conference on Emerging eLearning Technologies and Applications (ICETA 2008),

- Stará Lesná, The High Tatras, September 11-13, elfa, s.r.o., 2008, ISBN 978-80-8086-089-9
- [8] AMD: An Introduction do OpenCL [online], 2012, Available: <http://www.amd.com/us/products/technologies/stream-technology/opencvcl/Pages/opencvcl-intro.aspx>.
 - [9] Samuel Neves: Cryptography in GPUs, 2009, Departamento de Engenharia Informática, Universidade de Coimbra.
 - [10] Agosta, G.; Barengi, A.; DeSantis, F.; DiBiagio, A.; Pelosi, G.; , "Fast Disk Encryption through GPGPU Acceleration," Parallel and Distributed Computing, Applications and Technologies, 2009 International Conference on , vol., no., pp.102-109, 8-11 Dec. 2009.
 - [11] Brandon P. Luken, Ming Ouyang, and Ahmed H. Desoky: AES and DES Encryption with GP, 2009, Computer Engineering and Computer Science Department, University of Louisville.
 - [12] Owen Harrison and John Waldron: AES Encryption Implementation and Analysis on Commodity Graphics Processing Units, 2006, Computer Architecture Group, Trinity College, Dublin.
 - [13] Maksim Bobrov: Cryptographic Algorithm Acceleration Using CUDA Enabled GPUs in Typical System Configurations, 2010, Rochester Institute of Technology, Kate Gleason College of Engineering.
 - [14] Robert Szerwinski and Tim Guneyasu: Exploiting the Power of GPUs for Asymmetric Cryptography, 2009, Horst Gortz Institute for IT Security, Ruhr University Bochum, Germany.

Hamiltonian shortest path approximation optimizing a bitmap engraver tool path

¹Mišel BATMEND (2nd year), ²Martin MIKULÁK (bachelor student)
Supervisor: ³Daniela PERDUKOVÁ

^{1,2,3} Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

¹misel.batmend@tuke.sk, ²martin.mikulak@gmail.com, ³daniela.perdukova@tuke.sk

Abstract— This paper focuses on design and implementation of algorithm, which optimizes tool path of bitmap CNC engraver. We named the new algorithm Segmented line-by-line and it is a heuristic approximation of Hamiltonian shortest path. The algorithm searches for shortest path connecting all bitmap pixels that need to be engraved with respect to specific constraints given by hardware of bitmap engraver. New algorithm was implemented and tested on a real device. Results of experiments are listed. Concerning a set of 17 real-life bitmaps, the new Segmented line-by-line algorithm saves from 7% up to 17% of overall engraving time compared to original line-by-line tool path.

Keywords— engraving path, optimization, bitmap engraving

I. INTRODUCTION

In recent years, the bitmap engraving machine based on computer numerical control system has appeared on the market as a high science and technology product. Finding an optimal path for bitmap engraver tool can be treated as finding a solution to Traveler Salesman Problem (TSP) problem: Given a list of cities and their pair wise distances, the task is to find the shortest possible route that visits each city exactly once. The problem was first formulated as a mathematical problem in 1930 and is one of the most intensively studied problems in optimization [1]. Solution to TSP problem is called Hamiltonian shortest path. In theoretical computer science, a graph theory is often employed to model the TSP problem. The cities and routes are represented by vertices and edges in graph, respectively. The distance between cities is edge weight in a graph. A monochrome bitmap can be treated as a graph, with white pixels as vertices.

Image engraver is basically XY plotter engraving monochrome bitmap image on solid material by disrupting polished surface. It creates dots based on white pixels distribution in bitmap, thus building up an image out of thousands of dots, using a common halftoning technique to create an illusion of shades. The monochrome bitmap can be treated as weighted undirected graph, where weights are represented by spatial distances between pixels. In such scenario, there are tens of thousands vertices in a graph, where shortest Hamiltonian path needs to be found. This is, by now,

not feasible due to computational complexity of this problem. Thus an approximation of optimal path has to be found. Finding a shortest path is suitable to minimize the unnecessary tool movements, decreasing the time needed to engrave a picture.

Computing a solution to TSP problem involves generally two approaches: exact and heuristic algorithms. By now the best exact algorithms are running in time $O(2^n)$ [2]. Improving these time bounds seems to be difficult [3]. Therefore the number of vertices in a graph is limited to a few dozens, to get the path computed in a reasonable time. For example, in April 2006 an instance with 85,900 points was solved using a supercomputer. It would take over 136 years, if computed on a single core 500 MHz computer [4].

On the other hand, heuristic and approximation algorithms quickly yield a “good solution”, which is often very close to optimal solution. For example constructions based on a minimum spanning tree have an approximation to optimal ratio of 2. The Christofides [6] algorithm achieves a ratio of 1.5. Computation complexity of such algorithms is often very close to a linear time [5].

II. BITMAP ENGRAVER

Bitmap engraver is an XY plotter which uses a percussion diamond tool to engrave images into solid materials. As a model for engraving, monochrome bitmap image is used. White bitmap pixels are engraved. Engravers are often rather small, portable devices, which can be placed on top of engraved material. The material is in most cases a polished stone plate. Engravers are usually designed in a way that the Y axis has a bigger mass than axis X. This is because it carries the other axis actuator and mechanics. Thus the speed of movements in Y axis might be smaller than in X axis. Moreover moving Y axis causes more vibration, because of bigger mass. As a result, movements in X axis are preferred. Simultaneous movement of both axes is often not feasible. This means that the tool can only move in orthogonal directions, which is sufficient for bitmap engraving.

A. Engraver properties

For experimental purposes in this paper, we are using mb2300 bitmap engraver manufactured by Batmend. Its main

use is in engraving black granite. Its resolution is 180 dpi. Moving one pixel distance in X axis takes 2.6 ms, in Y axis 3.0 ms. Engraving a dot takes 22 ms. Normally engraved picture is approximately 850x1100 pixels big, contains about 200,000 dots and takes about 1h30min to engrave. The engraver is controlled by embedded microcontroller with rather small computing power (8bit, 20MHz). Therefore is the tool path pre-generated in a PC and fetched into microcontroller in suitable form.

B. Tool path

Commonly used tool path of bitmap engravers is generated line-by-line, from top to bottom of the bitmap. Such path is generated by computationally simple algorithm, but in most cases is far from optimal path. For example, in case of bitmap with two main clusters of dots, there is huge amount of unnecessary movements between those two clusters performed during engraving. In Fig.1, there is a small 10x4 pixel picture with two white pixel clusters. The tool path starts at "A" label and continues moving in rows. All white pixels are engraved. Unnecessary movements between clusters are evident.

The advantage of line-by-line path is that it uses as few Y axis movements as possible. For M x N image, the number of Y axis movements is M. Number of X axis movements depends on a wrapper of bitmap. In a worst case, when the bitmap wrapper is rectangular, the number of X axis movements is M x N.

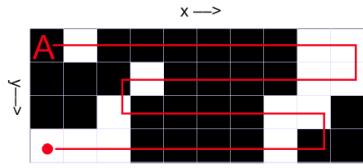


Figure 1. Line-by-line tool path

III. ALGORITHM

A. Requirements

There are two basic requirements on a new path finding algorithm. Firstly, movements in Y direction cost more than in X movements. The cost ratio might be as big as 1:1000 or even more. Secondly, algorithm must be very fast, that means that its runtime should not exceed tens of seconds. This constraint comes out of motivation, which is to shorten an overall engraving time. Overall time can be computed as a sum of subtask times:

$$t_{ALL} = t_{COMP} + t_{MOVE} + t_{DOT} \quad (1)$$

where t_{ALL} is overall engraving time, t_{COMP} is time of computing path trajectory, t_{MOVE} is duration of motor movements and t_{DOT} is time spent to engrave one dot (pixel). By shortening a tool path we can decrease motor movement time, since less movements are needed to complete the whole task. On the other hand, overall engraving time is also dependent on computing time of tool path algorithm. That requires a computing time not to exceed time saved on motor movements. Otherwise, the overall engraving time would not be shortened.

B. Description

The new segmented line-by-line algorithm is based on segmenting picture into smaller sections, which are engraved line-by-line. By this, we avoid moving over locations with no pixels to engrave.

The algorithm starts with making a binary mask with the same size as the picture has. At the beginning, whole mask is set to 1, representing wall in imaginary maze. Subsequently areas of size 13x9 pixels around each white pixel of original image are set to zero, which is a non-wall area in an imaginary maze.

The basic idea of pathfinder is following: Pathfinder gets to the top left pixel of non-wall area. Then it starts moving in rows, line-by-line, bouncing off the wall on both sides. When it gets to the bottom wall, it starts from the top left pixel again, seeing only pixels which have not already been engraved. This is done repeatedly, until the whole picture is engraved.

There are some minor refinements that shorten the tool path by connecting very small clusters of dots into bigger ones. The complete algorithm is described by a flow diagram in Fig.2.

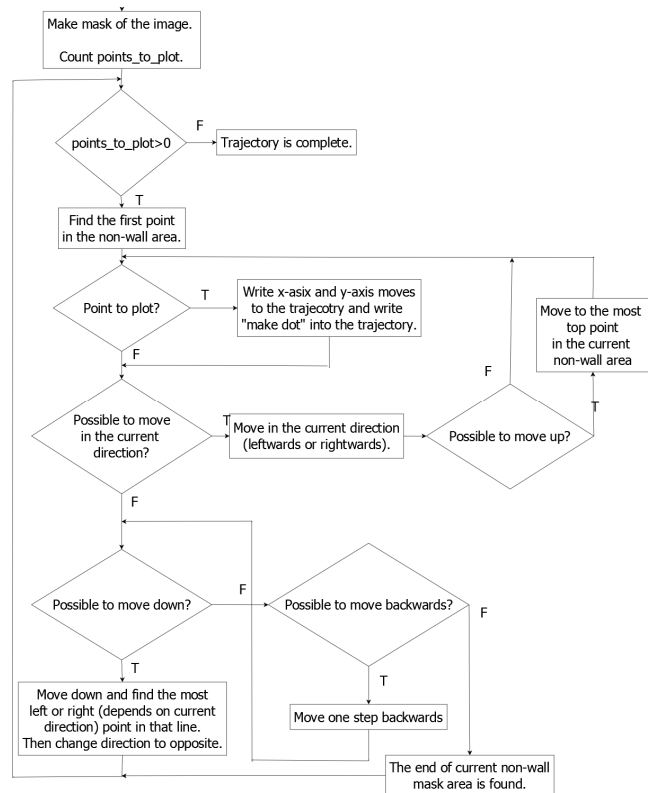


Figure 2. Flow diagram of segmented line-by-line algorithm

IV. RESULTS

The new algorithm was tested on 17 images which were taken from database of pictures normally engraved in industry. Correctness of generated trajectory was tested by software.

New trajectory of particular chosen image was fetched into engraver and the image was engraved. This experiment showed that there is no visual difference between old-trajectory picture and new-trajectory picture. Subsequently,

we compared new algorithm to original line-by-line algorithm. Moreover we compared it to a Minimum Spanning Tree (MST) based algorithm which we also programmed. The MST based algorithm meets the condition that the generated path is at most 2 times longer than shortest path. In our MST algorithm the weight of Y moves was equal to X moves weight. This made the MST generated path not suitable for practical use, since it generated many Y moves which led to excessive vibration of engraver. The compared factors of three algorithms were: 1. average computing time, 2. moves saved compared to line-by-line algorithm and 3. overall engraving time. Algorithms run on a normal PC with parameters: AMD Athlon64 X2 Dual core processor 4200+, 2.19 GHz, 896 MB RAM. To simplify the output table, we chose 3 representative images to put into a table. The first is a picture, where the new algorithm saved the least moves from all of tested pictures. The second one is a typical example and the last one is the picture where the new algorithm saved the most moves.

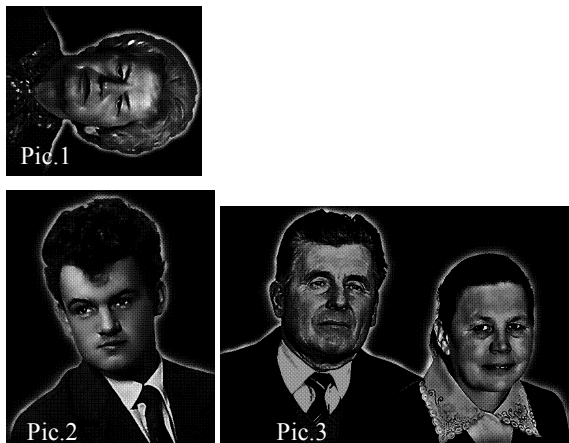


Figure 3. Pictures listed in Table I

TABLE I
ALGORITHM COMPARISON

Algorithm	Property	Pic.1 (12x10,5 cm)	Pic.2 (13x16 cm)	Pic.3 (22x15 cm)
Line-by-line	Average CPU time	0,89s	1,39s	2,40s
	Moves saved	0	0	0
	Engrav. time	0h:42m:28s	0h:55m:39s	1h:50m:37s
Segmented line-by-line	Average CPU time	3,98s	7,74s	8,98s
	Moves saved	17,85 %	36,99 %	41,86 %
	Engrav. time	0h:39m:24s	0h:46m:12s	1h:31m:20s
Minimum Spanning Tree	Average CPU time	0,58s	0,90s	1,73s
	Moves saved	47,12 %	57,04 %	60,37 %
	Engrav. time	0h:35m:05s	0h:41m:52s	1h:24m:17s

V. CONCLUSION

In a proposed paper, we designed and described the algorithm which optimizes tool path of bitmap engraving device. From theoretical point of view, it is approximation of Hamiltonian shortest path between image pixels. Because of engraver mechanical constraints, Y axis moves should be minimized, which was taken into account in new algorithm. Results from table I show, that new algorithm saved from 17

% to 41 % of tool movements, depending on a picture. This leads to saving 7 % to 17 % of overall engraving time. Computing time is a few seconds longer than original one. This difference is negligible compared to engraving time. Table I also shows superiority of MST based algorithm in both, computing time and path length. On the same pictures, it would be able to save 17 % to 23 % of engraving time. Unfortunately, MST generated path is not suitable for practical use, because of Y axis movements which cause excessive vibration while engraving. This was proven by experiment with MST generated path on real engraver. As a further research, an MST based algorithm with weight on Y axis moves could be implemented and tested.

REFERENCES

- [1] "Travelling salesman problem", From Wikipedia, the free encyclopedia, [online], cited 9.3.2012] http://en.wikipedia.org/wiki/Travelling_salesman_problem
- [2] Karp, R. M. : "Dynamic programming meets the principle of inclusion and exclusion", Operations Research Letters, Volume 1, Issue 2, April 1982, Pages 49-51, ISSN 0167-6377
- [3] Woeginger, G.J.: "Exact algorithms for NP-Hard problems: A survey", Combinatorial Optimization – Eureka, You Shrink! Lecture notes in computer science, vol. 2570, Springer, pp. 185–207, 2003
- [4] Applegate, D. L.; Bixby, R. M.; Chvátal, V.; Cook, W. J. , "The Traveling Salesman Problem", ISBN 0691129932, 2006
- [5] Chazelle, B.: "A minimum spanning tree algorithm with inverse-Ackermann type complexity", Journal of the Association for Computing Machinery 47 (6): 1028–1047, 2000
- [6] Christofides, N. : "Worst-case analysis of a new heuristic for the traveling salesman problem", Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, 1976.
- [7] Guang Dong; Xiaohan Zhang; Guangcai Cui: "Application of genetic algorithm on optimization of engraving paths", International Conference on Multimedia Technology 2011(ICMT), Pages 4850 - 4853, ISBN: 978-1-61284-771-9

How to apply recursion and corecursion in mathematical theory of programming

Pavol MACKO

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

pavol.macko@tuke.sk

Abstract—This paper deals with a problem of recursion and corecursion in computer science. It discusses about the morphisms, which use the recursion and corecursion. Work is specifically devoted to catamorphism, anamorphism and hylomorphism. In our previous papers we have presented our method of programming the recursive functions by using the hylomorphism. The main essence of this work is to describe how to apply our new method of programming in real systems. For application of our approach we have proposed a simple authentication system.

Keywords—anamorphism, catamorphism, hylomorphism, linear logic, Curry-Howard correspondence, authentication system, OCaml

I. INTRODUCTION

In our previous work [1] we showed and proved an alternative method for the factorial computation which is based on hylomorphism. Our idea is to use this method in real-life systems. We consider here very simple model of the authentication system where user is allowed to get in only if he/she is authorized to do it. There exist some other approaches of how to model a server implementation. In our method of authentication the user is not asked for input the password. The reason is that possible attacker in the network can catch the users' passwords. When only a sequence of seeming the random numbers (for attacker) through the network is sent, the attacker would not be able to recognize their meaning and to stole the enrolling data.

II. BASIC NOTIONS

We described our method for computation recursive functions, based on hylomorphism, by dual structures algebras and coalgebras. Usually we treat them in categories [2], [3]. We use special kind of algebras and coalgebras - an initial algebra and a final coalgebra, resp. It holds for initial algebra, that there exists the unique morphism from the initial algebra into any algebra. This morphism is called the *catamorphism*. By this morphism we described recursion. Dually, there exists final coalgebra, for which holds, that from any coalgebra exists unique morphism into final coalgebra, called *anamorphism* [4], by which we described corecursion. Composition of those morphisms is a new morphism which is called the *hylomorphism* [5], [6]. We applied it in the alternative way of the factorial computation [1], [7]. By using Curry-Howard correspondence [8] we made formulas in linear logic and then we proved it [1], what means that our program is correct and does not need verification. 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 [9]. While classical logic treats the sentences that are always true or false, in linear logic the truth value depends on an internal state of a dynamic system.

III. THE MODEL OF SYSTEM

Authentication system (Fig. 1) is based on the client-server process platform. The clients request from server an access to service. On the server side is a database containing the personal data of the users. Each user has the unique user identification number (*id*). When the client requests from server an access to service, he is confirmed whether the *id* is in database. If the user's *id* is in the database of users, then server generates two control numbers *m* and *n* by the following way:

$$m = 10 + \det \begin{pmatrix} cs(id) & cs(hour) \\ cs(min) & cs(sec) \end{pmatrix} \bmod 11$$

$$n = 21 + \det \begin{pmatrix} cs(id) & cs(hour) \\ cs(min) & cs(sec) \end{pmatrix} \bmod 10$$

where *cs* is the function of cipher sum, *det* is a determinant and *hour*, *min* and *sec* are the hour, minute and second of the time on the server when user requests an access to service from server. Here we have chosen the interval for the Fibonacci numbers from 10 to 20 for the first number and from 21 to 30 for the second number. This is because we would like to express numbers which are not very small for security and for usability of an algorithm also not very great. The numbers *m* and *n* can be sent to the user for instance by SMS message. User has his own simple application for calculating the response. After getting the data from server user inputs the numbers *m* and *n* to the user's application. The result for access is a string which contains *m*-th and *n*-th Fibonacci number. Both Fibonacci numbers (*Fib(m)* and *Fib(n)*) are converted to strings and the concatenated into one string. The same computation is made on the server side. After the client's application returns the result, the new generated string is sent to server. Then server compares the client's result and its eigen result. If the both strings are identical then user is granted to access.

We would like to remark that our model of system is only a proposition of an algorithm and we do not dwell on the data and transfer encryption.

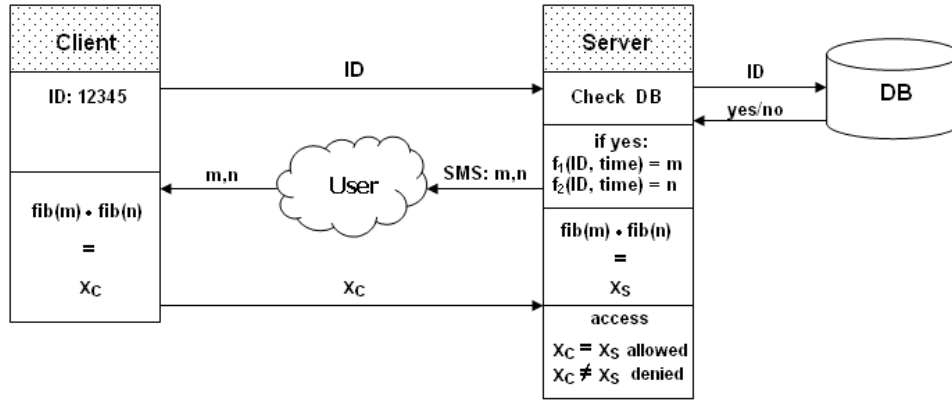


Fig. 1. The model of an authentication system

IV. HOW THE AUTHENTICATION PROCESS CALCULATES THE USER INPUT VALUES

The computation of the values sent to the user is based on mathematical structures introduced in previous chapters. The computation uses corecursive and recursive structures. An anamorphism is a generating structure and catamorphism is an eliminating structure. Firstly, the list of values is being generated by anamorphism and then the list is being consumed - eliminated by catamorphism. The elimination operation of catamorphism in that case is an addition.

For the computation of Fibonacci numbers we use the Pascal's triangle. We showed in [10] that it holds for the i -th Fibonacci number:

- 1) if the number i is even:

$$fib(i) = \sum_{\substack{k=0 \\ n=i-k-1}}^{\frac{i}{2}-1} \binom{n}{k} \quad (1)$$

- 2) if the number i is odd:

$$fib(i) = \sum_{\substack{k=0 \\ n=i-k-1}}^{\frac{i-1}{2}} \binom{n}{k} \quad (2)$$

For the combination number calculation we apply the previously defined factorial based on hylomorphism [11], [1].

V. THE IMPLEMENTATION

The implementation of function to calculate combination number in *OCaml* [12] is following:

```
let comb (n,k) = (fact (n)) / ((fact (n-k))
                             * (fact (k)) );;
```

The usage of the function $comb(n, k)$ is as usual:

```
# comb (3,2);;
- : int = 3
```

and the signature of the function is $int \times int \rightarrow int$.

A. The function $fib_ana()$

The function $fib_ana()$ is corecursive function, which is based on generating structure of anamorphism. This function corresponds to calculation of the n -th Fibonacci number by using formulas (1) and (2); its signature is

$int \rightarrow (int \times int)List$.

It generates the list of elements, where each element of the list is a tuple representing the combination number. Such a generated list is an input for the function $fib_cata()$. In the function $fib_ana()$ are used subsidiary functions $repeat_up()$, $repeat_down()$, and $zip()$.

```
let fib_ana i =
  if ((i mod 2) = 0) then
    zip (repeat_down (i/2) (i-1))
        (repeat_up (i/2) 0)
  else
    zip (repeat_down (((i-1)/2)+1) (i-1))
        (repeat_up (((i-1)/2)+1) 0);;
```

The usage of the function $fib_ana()$ is as follows:

```
# fib_ana 6;;
- : (int * int) list = [(5,0); (4,1); (3,2)]
```

and the result is the list of three tuples which represent the combination numbers.

The subsidiary function $repeat_up()$ is a simple recursive function which inserts n elements into the list. The numbers are generated increasingly starting from the value x .

```
let rec repeat_up n x =
  if n > 0
  then x::repeat_up (n-1) (x+1)
  else [];;

# repeat_up 3 0;;
- : int list = [0; 1; 2]
```

Similarly, the subsidiary function $repeat_down()$ inserts n elements into the list. The numbers are generated decreasingly starting from the value x .

```
let rec repeat_down n x =
  if n > 0
  then x::repeat_down (n-1) (x-1)
  else [];;

# repeat_down 3 5;;
- : int list = [5; 4; 3]
```


Finally, the function *zip()* is a recursive function. Its input are two lists: the first is a result of the function *repeat_up()* and the second is a result of *repeat_down()*. The result of the function *zip* is a new list which elements are tuples consisting of elements of both input list - the elements of input lists are "zipped" into a tuples and inserted into one list. The function has a signature

$$intList \times intList \rightarrow (int \times int)List.$$

```
let rec zip lst1 lst2 = match lst1, lst2 with
| [], _ -> []
| _, [] -> []
| (x :: xs), (y :: ys) -> (x, y) :: (zip xs ys);;

# zip [5; 4; 3] [0; 1; 2];;
- : (int * int) list = [(5, 0); (4, 1); (3, 2)]
```

B. The function *cata_fib()*

The function *cata_fib()* is a recursive function. Its input is a list of tuples of integers that represent the combination number from the Pascal's triangle. Function is based on the eliminating structure - the catamorphism. Input list is eliminated by calculating the combination number from the concrete tuple by using the function *comb()* and then adding all that values into one sum. After all the list is consumed, the function *fib_cata()* returns the *n*-th Fibonacci number. Our method does not need to calculate previous two Fibonacci numbers for the *n*-th Fibonacci numbers. The result is obtained just by adding the concrete combination numbers. The signature of the function *fib_cata()* is

$$(int \times int)List \rightarrow int.$$

```
let rec fib_cata list =
  match list with
  | [] -> 0
  | head :: tail -> comb (List.hd list) +
    (fib_cata tail);;

# fib_cata [(5, 0); (4, 1); (3, 2)];;
- : int = 8
```

C. The function *fib()*

Finally, the function *fib()* is a composition of the functions *fib_ana()* and *fib_cata()*. It generates the list of tuples which represent the combination numbers and then the list is eliminated by adding all the values into one sum. The result of the function *fib()* is the *n*-th combination number without knowing its previous two numbers in the sequence of the Fibonacci numbers. Signature of the function *fib()* is

$$int \rightarrow int.$$

```
let fib i =
  fib_cata (fib_ana i);;

# fib 6;;
- : int = 8
```

VI. CONCLUSION

In this contribution we presented how to use this new method of programming recursive functions in real-life systems. This method is based on the hylomorphism which it is the composition of catamorphism and anamorphism. We applied our approach in simple authentication system. The idea of our authentication system is that user is not requested to input the password but to make some calculations and then send those results to server. If the results have been calculated correctly, the user is granted to access. Our next goal is going to be the extension of this approach in other categorical structures based on recursive coalgebras and constructing of the appropriate description of the computation in linear logic and to describe the behavior of the authentication system by using categorical structures.

REFERENCES

- [1] P. Macko, "The expressing of structures for computation in linear logic," P. from 11th Scientific Conference of Young Researchers SCYR 2011, Ed., May 17th, 2011, pp. 377–380, Herľany, Slovakia. ISBN 978-80-553-0644-5.
- [2] M. Barr and C. Wells, *Category Theory for Computing Science*. Prentice Hall International, 1990, ISBN 0-13-120486-6.
- [3] J. J. y. . . v. . . n. . . p. . . n. . I. V. Slodičák, title =Some useful structures for categorical approach for program behavior.
- [4] J. Adámek, D. Lücke, and S. Milius, "Recursive coalgebras of finitary functors," *ITA*, vol. 41, no. 4, pp. 447–462, 2007.
- [5] M. Fokkinga and E. Meijer, "Program calculation properties of continuous algebras," CWI, Amsterdam, Tech. Rep., 1991, CS-R9104.
- [6] T. Uustalu, V. Vene, and V. Capretta, "Corecursive algebras: a study of general corecursion," H. Springer, Ed., 2009, Berlin, ISBN 978-3-642-10451-0.
- [7] V. Slodičák and P. Macko, "The rôle of linear logic in coalgebraical approach of computing," *Journal of Information and Organizational Sciences*, vol. 35, no. 2, pp. 197–213, 2011, ISBN 0-13-120486-6.
- [8] M. Sørensen and P. Urzyczyn, *Lectures on the Curry-Howard Isomorphism*, University of Copenhagen and University of Warsaw, 1999.
- [9] J.-Y. Girard, "Linear logic," *Theoretical Computer Science*, vol. 50, no. 1, pp. 1–102, 1987, ISSN 0304-3975.
- [10] V. Slodičák and P. Macko, "New approaches in functional programming using algebras and coalgebras," in *European Joint Conferences on Theory and Practice of Software - ETAPS 2011*. Universität des Saarlandes, Saarbrücken, Germany, March 2011, pp. 13–23, ISBN 978-963-284-188-5.
- [11] V. Slodičák and P. Macko, "How to apply linear logic in coalgebraical approach of computing," in *CECIS 2011: proceedings of the 22nd Central European Conference on Information and Intelligent Systems*. Varaždin, University of Zagreb, Croatia., September 21st-23rd 2011, ISSN 1847-2001.
- [12] E. Chailoux, P. Manoury, and B. Pagano, *Developing Applications With objective Caml*. O'REILLY & Associates, Paris, ISBN 2-84177-121-0, 2000.

Channel parameters simulation using NS-2 and MyEvalvid-NT environments

¹Ján VALISKA (1st year), ²Ondrej KOVÁČ (1nd year)

Supervisor: ³Stanislav MARCHEVSKÝ

^{1,2,3}Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

¹jan.valiska@tuke.sk, ²ondrej.kovac@tuke.sk, ²stanislav.marchevsky@tuke.sk

Abstract—This article deals with modeling the transmission channels using usual channel models, the transmission of video sequences over the canals and evaluation using basic metrics for assessing video quality. For building and simulation of the channels is used network simulator NS-2. To be able to transfer through the channel and evaluate the video sequences, it was necessary to use an extension for NS-2 called MyEvalvid2 that also serves for PSNR and SSIM metrics computation and visualization of measurement data. During the measurements were used various parameters of error channel models, namely the Gilbert-Eliot and Binary Symmetric model, which are used in the NS-2 simulation environment scenario.

Keywords—channel modeling, network simulator, packet loss, evalvid

I. INTRODUCTION

NS-2 is an open-source event-driven simulator designed specifically for research in computer communication networks. Since its inception in 1989, NS2 has continuously gained tremendous interest from industry, academia, and government. Having been under constant investigation and enhancement for years, NS-2 now contains modules for numerous network components such as routing, transport layer protocol, application, etc. To investigate network performance, researchers can simply use an easy-to-use scripting language to configure a network, and observe results generated by NS-2. Undoubtedly, NS-2 has become the most widely used open source network simulator [1].

Generally, network simulators facilitate the study of network architectures such as IP QoS implementations. NS-2 is one of them. Network simulation scripts in NS-2 are used to create the network scenarios and upon the completion of the simulation, trace files that capture events occurring in the network are produced [2]. The trace files would capture information that could be used in performance study, e.g. the amount of packets transferred from source to destination, the delay in packets, packet loss etc [3]. However, the trace file is just a block of ASCII data in a file and quite cumbersome to access using some form of post-processing technique [4].

A. Software organization

NS-2 is an object oriented simulator written in C++ that uses OTcl (an object oriented version of Tcl) as the command and configuration interface. Thus, the entire software hierarchy is written in C++, with OTcl used as a front end. C++ code serves as a backbone for the whole simulation process. NS-2 utilizes

two languages as far as simulator has two different kinds of things it needs to do. On the one hand, detailed simulations of protocols require a system programming language which can efficiently manipulate bytes, packet headers, and implement algorithms that run over large data sets. For these tasks, run-time speed is important and turn-around time (run simulation, find bug, fix bug, recompile, re-run) is less important. On the other hand, a large part of network research involves slightly varying parameters or configurations, or quickly exploring a number of scenarios. In these cases, iteration time (change the model and re-run) is more important. Since configuration runs once (at the beginning of the simulation), runtime of this part of the task is less important [5]. One drawback of combining two languages is that debugging becomes more complicated than with one language alone.

The basic simulator objects that NS-2 includes are Nodes, Links, Agents and Applications. Nodes and Links define the network topology. Agents represent endpoints where network-layer packets are constructed or consumed and are commonly used to implement protocols at various levels. Applications are sources that send and receive data. The objects are connected to each other in a layered fashion as shows Fig. 1 [5].

There are two major types of agents in NS-2: one-way and two-way agents. The one-way agent consists of one sender and one corresponding receiver. Data can only be sent in one direction and often connection establishment/teardown is not performed. The two-way agent is both - a sender and receiver, supporting bi-directional data transfers and could be very well a real-world counterpart. Currently, NS-2 offers support for various versions of TCP and for TRFC.

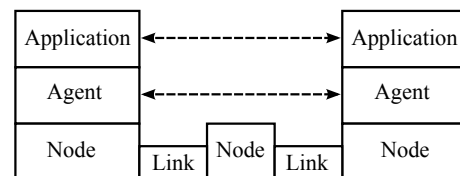


Fig. 1: Main NS-2 objects and their interconnection.

II. PACKET LOSS MODELING

Communication channel can be classified into two categories: memoryless channels and channels with memory [6]. Space or satellite channels are a few examples of memoryless channels. On these channels, a received signal at a given time interval depends only on the transmitted signal in that interval, but not on previous signal transmissions. As a result, errors

appear randomly on memoryless channels, and therefore these channels are often referred as random-error channels [7]. Many real communication channels, however, utilize memory. Errors on these channels tend to occur in clusters or bursts. This bursty behavior is common in mobile communication channels, telephone channels, and magnetic or optical recording systems. In simulation section, we focused on two simple channel models, the binary symmetric channel (BSC) model representing a memoryless channel, and the Gilbert-Elliott channel model representing a channel with memory.

A. Gilbert model

Packet loss measurements on the Internet have shown that the probability of loss episodes of length k decreases approximately geometrically with increase of k [8]. Thus, it is possible to use simpler packet loss model, e.g. Gilbert model. Gilbert model is a special case of k -th order Markov chain model, see Fig. 2.

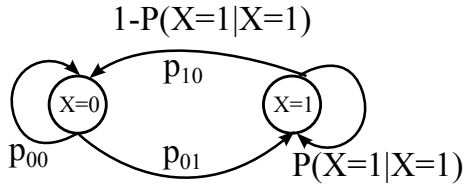


Fig. 2: Gilbert model.

In this model, 0 represents state with no packet loss and 1 represents the state of packet being lost. The matrix for transition probabilities and for state probabilities can be expressed in form:

$$\begin{bmatrix} 1 - p_{01} & p_{10} \\ p_{01} & 1 - p_{10} \end{bmatrix} \begin{bmatrix} P(X=0) \\ P(X=1) \end{bmatrix} = \begin{bmatrix} P(X=0) \\ P(X=1) \end{bmatrix} \quad (1)$$

For unconditional probability $P(X=1)$ holds the following equation:

$$P(X=1) = \frac{p_{01}}{p_{01} + p_{10}} \quad (2)$$

If previous packet is lost, then for conditional probability of having loss holds:

$$P(X=1|X=1) = 1 - p_{10} \quad (3)$$

Gilbert model memorizes only the previous state, thus the probability, that the next packet will be lost is dependent only on the previous state [9].

Transition probabilities p_{01} and p_{10} can be expressed with the following equations:

$$p_{01} = P(X=1|X=0) = \sum_{k=1}^{\infty} \frac{o_k}{a} \quad (4)$$

$$1 - p_{10} = P(X=1|X=1) = \frac{\sum_{k=1}^{\infty} (k-1)o_k}{d-1} \quad (5)$$

The probability of having a lost episode with length k [8]:

$$p_k = (1 - p_{10})^{k-1} p_{10} \quad (6)$$

B. The Binary Symmetric channel model

The binary symmetric channel model (BSC) is a simple model for discrete memoryless channels (DMC). A discrete memoryless channel of M -ary inputs and Q -ary outputs is defined by the transmission probabilities $P(j/i)$, where $0 \leq i \leq (M-1)$ and $0 \leq j \leq (Q-1)$.

In the BSC model, $M=2$ and $Q=2$. The model is completely defined by a channel error rate or crossover probability ϵ as shown in Fig. 3 [8].

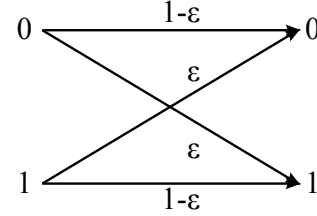


Fig. 3: The Binary symmetric channel model.

III. EVALVID-VIDEO QUALITY EVALUATION TOOLSET

EvalVid is a framework and toolset for quality evaluation of the video transmitted over a real or simulated communication network. It is targeted for researchers who want to evaluate their network designs or setups in terms of user perceived video quality [10].

Besides measuring QoS parameters of the underlying network, like loss rates, delays, and jitter, standard video quality metrics like PSNR and SSIM [11] and a subjective video quality evaluation metric of the received video are provided. Currently, H.264 MPEG-4 and H.263 are supported. AAC support is also included, though the perceptual quality evaluation has to be done by external tools implementing dedicated metrics like PESQ or PEAQ [10].

A. MyEvalvid-NT

A novel realistic simulation tool-set known as MyEvalvid-NT for evaluating video delivery over simulated networks was created. This tool-set integrates Evalvid and NS2 [12]. The toolset can adopt traffic trace file from the output of encoding raw YUV video or from Internet traffic traces. Fig. 4 shows the system architecture of MyEvalvid-NT [8]. In this framework we can use two different input sources: publicly available traffic trace and the trace from encoding raw YUV video. Three connecting interfaces, namely MyTraffic-Trace2, MyUDP, and MyEvalvid Sink2, are implemented in NS2 environment [10].

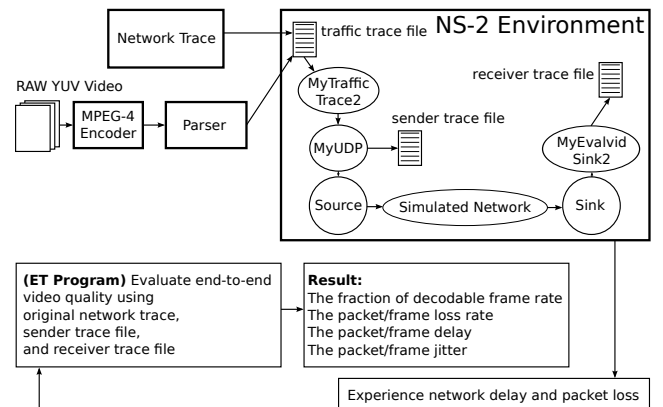


Fig. 4: The system architecture of MyEvalvid-NT

IV. SIMULATIONS

In General, *PSNR* is the metric used to measure the perceived quality at the end user. However, *PSNR* takes a long time to compare each pixel in the original frame and distorted frame to get *PSNR* value. Alternatively, we can use decodable frame rate (*Q*) [10] to measure video quality at the receiver, which takes less time compared to *PSNR*. Therefore, in this paper, we use decodable frame rate as one of our main video evaluation metric. Through simulations, we also reveal that decodable frame rate can evaluate the perceived quality well by an end user.

A. Used topology

The simulation topology used in this paper is shown in Fig. 5. The video server transmits video streams over the Internet and wireless links to the video receiver. The maximum transmission packet size is 1024 bytes. The link between the wireless access point and the video receiver is IEEE 802.11b 11Mbps. For simplicity, we assume that the link between the video server and the wireless access point has a 10Mbps bandwidth and 10ms latency [10].

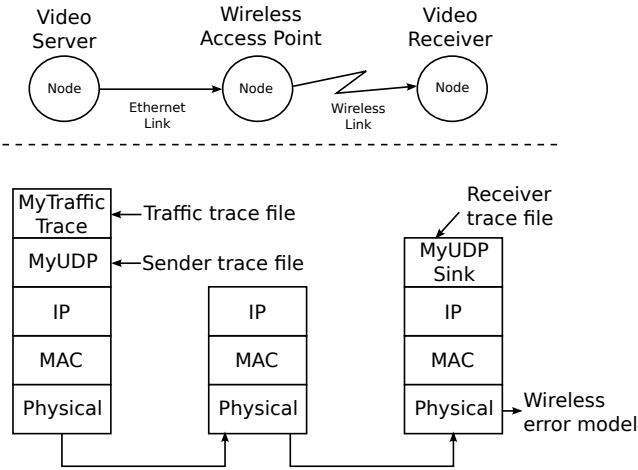


Fig. 5: Simulation topology

B. Simulation results

The several simulations were done due to evaluation of used scenario. The results were processed in graphs and tables. In each simulation, the same values, 10Mbps of bit rate and maximal delay 10ms were used.

1) *Simulation results using BSC multicast channel model:* The first simulated model was BSC channel model in multicast wireless network [13]. In order to portray the graphs clearly, we simulated three different values of error rate: 1%, 10% and 50%. At the highest value of error rate - 50%, the *PSNR* value was less than 20dB. On the basis of further simulations, we compared the values *Q* and *PSNR*, that are listed in Table I. As is possible to see, the lower value of *Q*, the lower *PSNR* value and the lower quality of received video sequence. In case of $Q = 92,83\%$, the *PSNR* value was equal to 32.7dB, whereas in case of $Q = 7.88\%$, the *PSNR* value was equal to 12.3dB.

2) *Simulation results using BSC unicast channel model:* The second simulated model was BSC channel model in unicast wireless network [13]. Again, in order to portray the graphs clearly, we simulated three different values of error rate. Since there was no difference in results between error

rate values from 1% to 20%, we portrayed the graph for three values different from those in previous channel model, namely 1%, 30% and 80%. At the highest value of error rate - 80%, the *PSNR* value was less than 20dB, only in some frames *PSNR* exceeded 20dB value. Again, we compared the values of *Q* and *PSNR*, that are present in Table II. As we already know, the lower value of *Q*, the lower *PSNR* value and the lower quality of received video sequence. In case of $Q = 95,69\%$, the *PSNR* value was equal to 32.49dB, what we consider as quite good video quality.

3) *Simulation results using Gilbert-Elliott multicast channel model:* The last simulated model was BSC channel model in unicast wireless network [13]. We simulated three different values of error rate: 1%, 10% and 50%, as it was in first simulated channel model. As is shown in Fig. 7, the first two values of error rate achieved approximately the same *PSNR* values. They were different only in some cases during more dynamic moments in video sequence. At the highest value of error rate - 50%, the *PSNR* value was less than 20dB, only from frame no. 120 to frame no. 160, *PSNR* exceeds 25dB value.

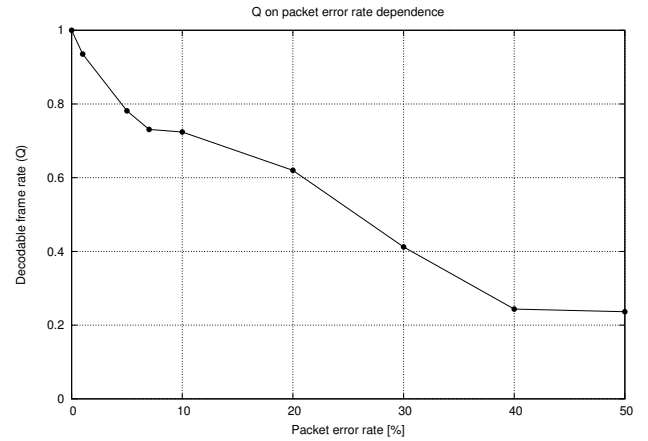


Fig. 6: The packet error rate in % and *Q* values comparison for Gilbert-Elliott multicast channel model

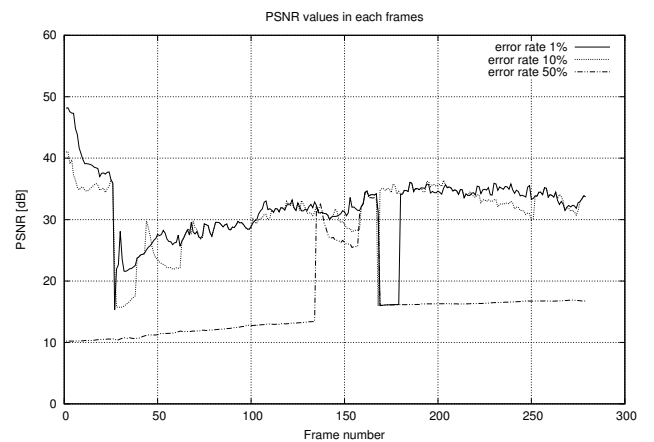


Fig. 7: *PSNR* values for Gilbert-Elliott multicast channel model

REFERENCES

TABLE I: PSNR and Q values for BSC multicast channel model

Errorrate	PSNR [dB]	Q
0.01	32.07	0.9283
0.05	30.81	0.8566
0.1	29.16	0.7562
0.3	19.96	0.2974
0.5	12.301	0.0788

TABLE II: PSNR and Q values for BSC unicast channel model

Errorrate	PSNR [dB]	Q
0.01	34.97	1
0.1	34.97	1
0.3	32.49	0.9569
0.5	29.81	0.6774
0.6	15.98	0.1218

TABLE III: PSNR and Q values for BSC unicast channel model

Errorrate	PSNR [dB]	Q
0.01	31.40	0.9354
0.05	31.28	0.7813
0.1	30.86	0.7240
0.3	25.52	0.4121
0.5	15.76	0.2365

As shows Table III, Q and $PSNR$ values are changing regularly, but at the level of error rate 30%, both values changed markedly, what we consider as worse video quality. Fig. 6 shows relation between Q and error rate values using Gilbert-Elliott multicast channel model.

V. CONCLUSION

In our paper, we focused on testing of new simulation tool-set MyEvalvid-NT, that connects Evalvid and NS-2 simulators. The benefit of such simulations is the possibility to test a real behavior of wireless or other networks and various variants offered by this tool-set. From the simulation results we learnt that the video quality of the lower packet error rate is superior to the higher packet error rate and corresponds to the results of decodable frame rate Q . In addition, since $PSNR$ is a well-known performance metric of evaluating video delivery quality, we also revealed that the Q metric reflects the behavior of $PSNR$ well [10]. Therefore we can evaluate the quality of delivered video more quickly through the tool-set, when we verify designs in multimedia communication.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

- [1] T. Issariyakul and E. Hossain, "Introduction to network simulator NS-2," *Springer Science +Business Media*, 2009, 233 Spring Street, New York, USA.
- [2] A. Salleh, Z. Ishak, N. Din, and Z. Jamaludin, "Trace analyzer for NS-2," *4-th Student Conference on Research and Development*, pp. 27–28, June 2006, Shah Alam, Selangor, Malaysia.
- [3] M. Uhrina, J. Hlubik, M. Bojmir, and M. Vaculik, "The impact of GOP size and packet loss on the video quality," *13-th International Conference on Research in Telecommunication Technologies – RTT 2011*, 2011, techov, Czech Republic.
- [4] A. Huszak and S. Imre, "Analysing GOP structure and packet loss effects on error propagation in MPEG-4 video streams," *4-th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pp. 1–5, 2010.
- [5] N. Mattsson, "A DCCP module for NS-2," Master's thesis, Lulea University of Technology, 2004.
- [6] P. Peda, J. Ethridge, M. Baines, and F. Shallwani, "A network simulator differentiated services implementation," *Open IP, Nortel Networks*, July 2000.
- [7] H. Sanneck, G. Carle, and R. Koodli, "A framework model for packet loss metrics based on loss run-lengths," *Proceedings of the SPIE/ACM SIGMM Multimedia Computing and Networking Conference 2000*, pp. 177–187, 2000, San Jose (California, USA).
- [8] V. Markovski, F. Xue, and L. Trajkovic, "Simulation and analysis of packet loss in video transfers using user datagram protocol," *The Journal of Supercomputing*, Kluwer Academic Publishers, vol. 20, pp. 175–196, 2001, Hingham (Massachusetts, USA).
- [9] R. Polec and T. Karubíková, *Stochastické modely v telekomunikáciách I*. Fond Jozefa Murgaša pre telekomunikácie n.f. vo vydavateľstve FABER, May 1999, vol. 1.
- [10] Y. Chia-YU, K. Chih-Heng, S. Ce-Kuen, and N. Chilamkurti, "My evalvid-NT-A simulation toolset for video transmission and quality evaluation," *TENCON 2006, IEEE Region 10 Conference*, pp. 1–4, November 2006, Hong-Kong.
- [11] P. Poca, R. Hudec, and M. Bojmir, "Performance evaluation of video transmission over 802.11b WLANs from the video quality perspective," *The Mediterranean Journal of Computers and Networks*, vol. 5, no. 2, pp. 53–58, 2009.
- [12] E. Altman and T. Jimenez, "NS simulator for beginners," *Lecture notes, 2003-2004*, December 2003, Univ. De Los Andes, Mérida, Venezuela and ESSI, Sophia-Antipolis, France.
- [13] P. Adamčin, "Maskovanie chýb pri prenose videotokov bezdrôtovým prostredím v dôsledku preťaženia siete v programovom prostredí NS-2," Master's thesis, Technical university of Kosice, May 2011.

Ideal MPM-20 jet engine running process with zero degrees of freedom

¹Vladimír GAŠPAR (1st year), ²Tomáš KAROL (3rd year) and ³Cecília HAVRILOVÁ
Supervisor: ³Ladislav MADARÁSZ

^{1,2,3}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹vladimir.gaspar@tuke.sk, ²tomas.karol@tuke.sk, ³cecilia.havrilova@student.tuke.sk

Abstract—This paper presents the basic data analysis of experimental data obtained from the Laboratory of Intelligent Control Systems of Jet Engines. There are two different aims of this paper. The first aim is to propose a theoretically ideal graph of the engine running process. The second aim is to evaluate whether there is any notable wear level.

Keywords—ideal running process, MPM-20, data analysis, error correction, wear level.

I. INTRODUCTION

Data analysis for safety purposes and atypical states detection is very crucial when considering a complex system. In complex thermodynamic systems the safety concerns are even greater. This is why there is a need for creating an ideal running process of such a system. The object of our research is a converted turbo-starter jet engine TS-21 labeled as MPM-20 (small turbojet engine, Fig. 1).

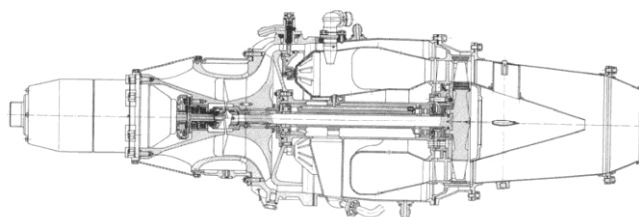


Fig. 1. Wire model of the converted turbojet engine MPM-20

Since these starter engines are used only for a short period of time (until the main engine is at idle RPM) the operation time of the MPM-20 is limited to 1-2 minutes [1]. After this time period is exceeded, the wear of the engine may increase rapidly. Lowering the wear level requires evaluating the ideal running process of the turbojet engine which will be a theoretical mean for the real engine operation.

In real-time operation, the engine should oscillate around the proposed mean values. We also have to choose suitable attributes that have the greatest influence on the engine operation and logically divide the engine operation into discrete operation groups. These operation groups will be chosen according to the startup, operation and fall time of the engine. For this experiment we consider 0 degrees of freedom. This means that the engine will be running with its

default settings without any digital feedback control.

II. EXPERIMENT SETTINGS

A. Expectations

We expect that every measurement will have the same behavior for each measured attributes. The main differences may be caused by different date of the measuring (ambient environment temperature), the order of measurements (engine is preheated by previous launches), other external forces and possibly the wear level of the engine parts. Time shift in each measured attribute may be caused by late start of measuring or, in some cases, the length of the engine idle operation period [3].

B. Fuel intake settings

The fuel intake is managed by hydro-mechanical elements (screws and pressure valves). The expected fuel intake signal is depicted in the Fig. 2.

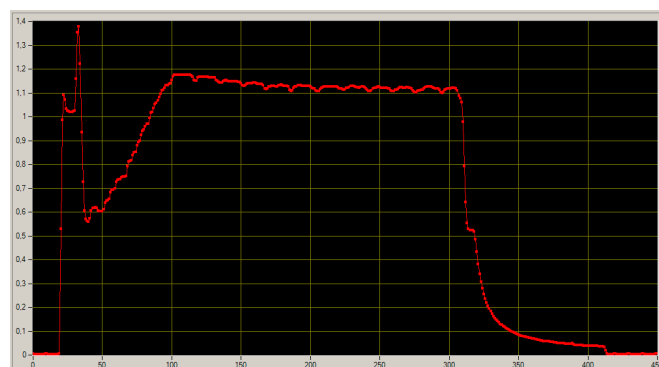


Fig. 2. Expected engine fuel intake (x-axis: time samples ; y-axis: fuel intake in liters per minute)

C. Carrying out experiments

Experiments from which the measured data had been obtained were carried out using a normal (hot) startup mode. It means that the operation process [1] [2] [4]:

1. Starts with electromotor from 0 rpm to approximately 13 500 rpm. Simultaneously the jet fuel igniter plugs, fuel and oil pump are turned on.
2. When the pressure at the pressure switch is over 0,02-0,045kPa, the electromagnetic valve opens and the fuel injection and oil flow is enabled.

3. If the fuel is ignited and the speed of compressor exceeds 20 000 rpm the electromotor and jet fuel igniter plugs are disconnected.
4. When the pressure exceeds 0,19-0,21kPa, the pressure switch opens the bypass electromagnetic valve (EMV).
5. The engine reaches idle state (speed) when the compressor rpm is approximately between 40 000-48 000 rpm and the fuel intake is stable. The maximum idle rpm is fully dependent on the fuel intake. Measurements have been carried out at approximately 1,2 liters per minute fuel intake in the idle engine state.
6. When the main engine switch is switched off the pressure decreases, the bypass is closed and also the electromotor is fully disconnected so that it would not start automatically when the pressure at the electro magnetic switch decreases to specific limit.

III. DATA PROCESSING

To understand the data we firstly needed to know the distribution, variability and behavior of each attribute in time and relationships between attributes. For this purpose we have been using SPSS Clementine 11.1 (Fig. 3).

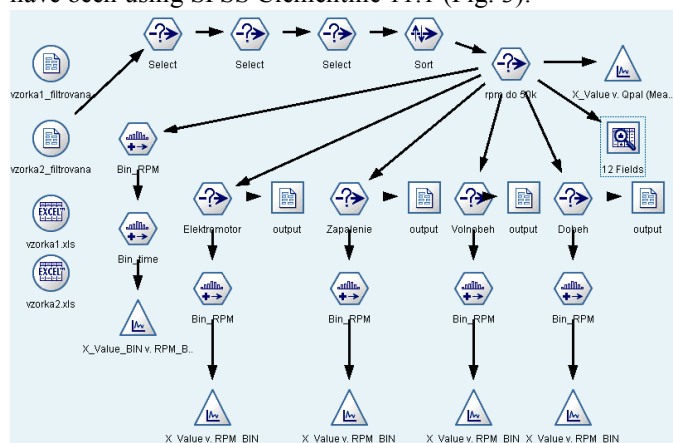


Fig. 4. Data processing diagrams created in SPSS Clementine

For comparing each measurement in time with normalized values of various attributes, we have created an analyzer program using .NET Framework and National Instruments Measurement Studio components for Visual Studio 2010.

A. Measurements selection

The ability to consider changes in wear level can only be done by comparing two different date ranges. This is why we divided measurements into two volumes. The first volume (9.2.2010 - 20.5.2010) contained 71 independent measurements and the second volume (11.7.2011-17.2.2012) contained 64 measurements, both before the measurement selection process.

In the next step, we removed measurements that did not cover the entire engine launch process since startup, until shutdown and 0 rpm of the compressor. In some cases unsuccessful (non ignited hot engine launches) or cold engine launches had been archived. These measurements cannot be used in ideal hot engine operation evaluation.

After this process of measurements selection, remaining files for analysis are:

- volume 1 - 41 measurements
- volume 2 - 40 measurements

B. Erroneous, extreme values processing and column reduction

Since there is a possibility that sensor errors may occur, we have to treat extreme values. The greatest number of errors has been observed in values received from the RPM sensor (Fig. 4) and thermocouple T_{3C} . The RPM sensor is reaching its limits and is very sensitive to environment changes (i.e. smoke, oil smudges and vibrations). Also the thermocouple T_{3C} reaches its limits because it measures the actual temperature in the combustion chamber of the engine. This temperature may reach up to 1300 °C. The T_{3C} thermocouple has been replaced over the measured period, so it is not suitable to be used in this analysis.



Fig. 4. Extreme erroneous value measured by the RPM sensor (x axis- time samples ; y axis - RPM)

C. Logical data division and grouping

The main data division has been carried out using SELECT nodes in SPSS Clementine. These nodes have been set to filter data from every independent situation. All properties of created groups are shown in the following table.

Tab. 1. Created data groups

	RPM	Time
Electromotor	< 14 000	----
Burning launch	>14 000 < 41 300	< 15 sec
Idle operation	> 41 300 < 50 000	> 15 sec < 50 sec
Shutdown and fall	< 41 300	> 50 sec

These groups have been selected according to specific differences between each state (see. Chapter II C) of the engine operation. Time samples for group division have been chosen by observations of the inflex points. The main changes in the engine dynamics and change of the operating situations may be observed in "RPM in time" graphs in form of these inflex points. They can be most significantly observed in measurements, when the engines was cold (Fig. 5.).

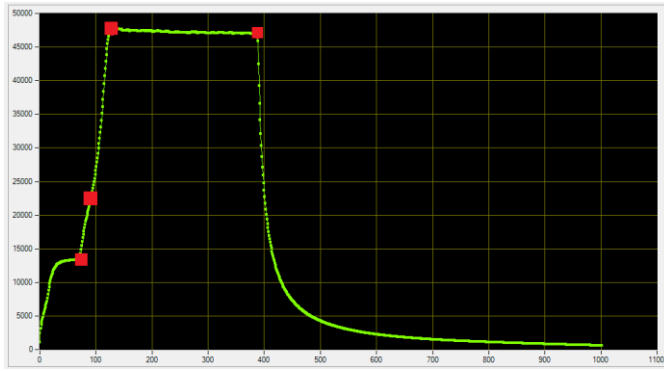


Fig. 5. Change of engine dynamics in form of inflex points (shown as red points)

Despite one more inflex point also occurs at approximately 20 000 RPM (bypass EMV opened and automatic combustion starts) there is no need to involve this situation into analysis. This inflex point may be very promising for analyzing failed starts or other erroneous engine operation states in the future (i.e. engine misfire, alternative fuels combustion and other combustion related problems).

After the selection and correction of chosen samples, we have trimmed the output graphs, so that the mean of the values distribution in time would be seen. These trimmed graphs of both samples can be observed in Fig. 6. and Fig. 7. If we apply the same trimming conditions on both samples, then after the end of processing we receive fully comparable samples of data. As we have already mentioned the main difference between these two graphs is the engine shut down process because the time of the engine idle state was not strictly measured when executing engine launches.

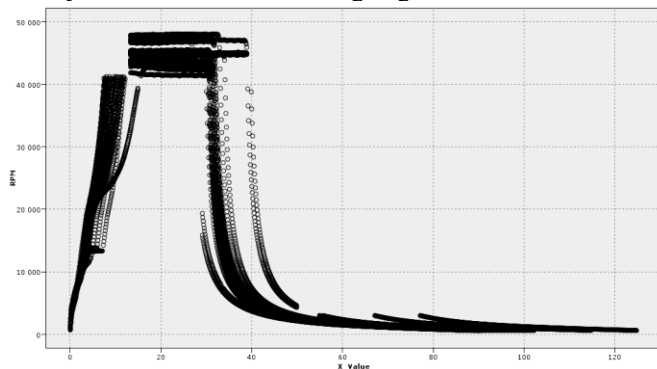


Fig. 6. Trimmed graph of sample volume 1

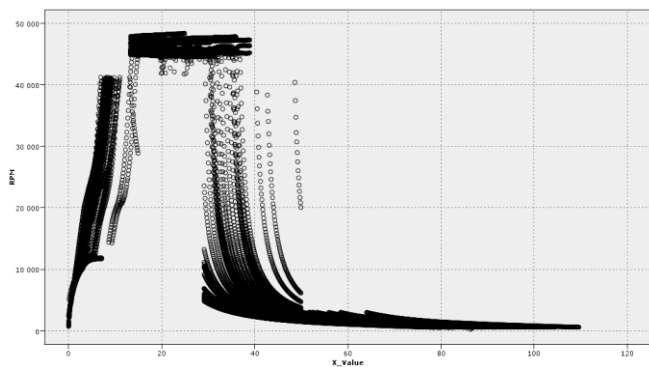


Fig. 7. Trimmed graph of sample volume 2

In dependence with the trimmed RPM every other attribute with high correlation with the RPM (pressures, output temperatures, fuel intake quantity see Fig. 8.) will be trimmed with respect to the chosen RPM and time groups.

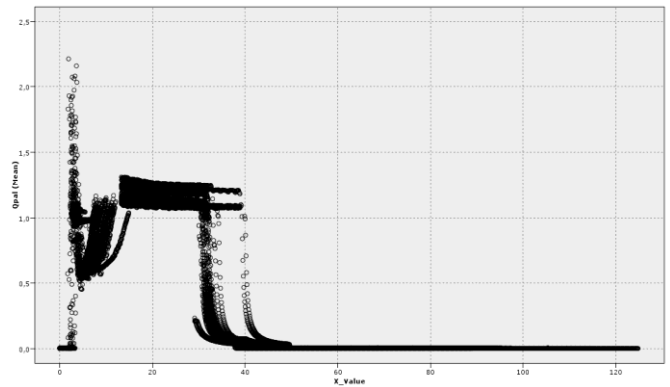


Fig. 8. Fuel intake in time according to trimmed RPM

IV. FINDINGS AND OUTCOMES

Ideal running process of the engine should comply with following facts:

- Engine should reach idle speed as soon as possible,
- engine should ignite the fuel in time when the fuel injection starts,
- at idle, engine speed should be constant if thrust (output nozzle diameter), fuel intake and pressure is constant,
- engine shutdown should have exponential decay character.

Based on these criteria and processed data we propose process graph depicted in Fig. 9. To slim up partial graphs of both chosen volumes and to point out all statistically important value densities we have used the binning method for RPM and time attributes. The mean curves of both volumes (graphs) and the intersection graph are also shown in Fig. 9.

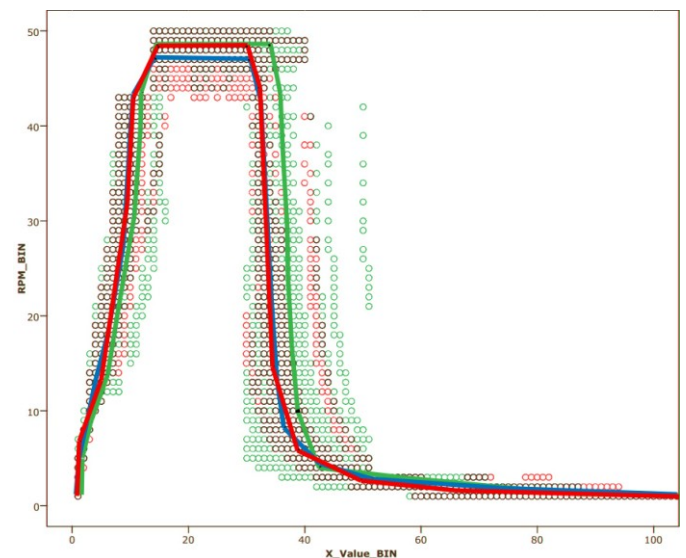


Fig. 9. Binned data intersections and their mean curves (points: green-volume2, red - volume1, brown - intersection ; curves: green - volume2, blue - volume1, red-intersection)

From this graph we may see that the differences between two chosen data ranges (volumes) are not significant and may only be caused by statistical error. However, when considering wear level, the mean startup part of the volume 2 curve is 5 time samples later than the startup part of the mean volume 1 curve.

Using analogy we may generalize outcomes of the previous graph into operation characteristics of other attributes, i.e. temperatures, pressures, tension (thrust), etc. As an example of the relations between P_{2c} , Q_{pal} and RPM we also present the startup graph (Fig. 9.) with respect to ideal RPM curve.

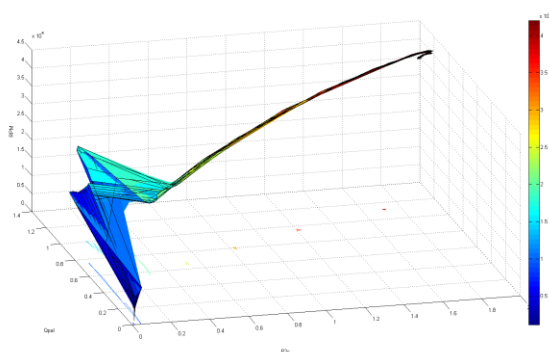


Fig. 9. Pressure and fuel intake relation with ideal RPM (x - fuel intake Q_{pal} , z - pressure P_{2c} , y - ideal RPM, color gradient - rpm level)

V. CONCLUSION

According to experimental analysis and data processing, we have proposed an ideal theoretical running process of the MPM-20 engine. This model may be used as a template for engine running and for discovery of atypical engine operation situations with 0 degrees of freedom when implemented real-time. If more degrees of freedom will be added to engine control process (fuel intake real-time control, thrust control), at least the startup and shutdown parts of the process will be identical.

We can also sum up the fact that the engine wear level is minimal and can be observed in the engine RPM graphs only in minimal extent when comparing two selected volumes.

ACKNOWLEDGMENT

The work presented in this paper was supported by VEGA, Grant Agency of Ministry of Education and Academy of Science of Slovak Republic under Grant No. 1/0298/12 – “Digital control of complex systems with two degrees of freedom”. The work presented in this paper was also supported by KEGA under Grant No. 018TUKE-4/2012 - Progressive methods of education in the area of control and modeling of complex systems object oriented on aircraft turbo-compressor engines. This support is very gratefully acknowledged.

REFERENCES

- [1] T. Lazár, L. Madarász (Eds.): Inovative Outputs from the Transformed Experimental Laboratory with a Small Turbojet Engine. elfa, s.r.o. Košice, 348pp. ISBN 978-80-8086-170-4 (2011).
- [2] L. Madarász, R. Andoga, L. Fözö: 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>, pp. 17 -38. ISBN: 978-953-307-213-5, (2010).
- [3] L. Madarász, R. Andoga, M. Českovič, M. Laššák, M. Šmelko, L. Fözö: Advanced approaches in modeling of complex electro-mechanical systems,

12th IEEE International Symposium on Computational Intelligence and Informatics, 21. - 22. November, Budapest, Hungary, ISBN 978-1-4577-0043-9, pp. 541 – 544 (2011).

- [4] T. Karol, V. Gašpar, L. Madarász, R. Andoga, L. Fözö, J. Judičák: Advanced approaches Artificial intelligence in diagnostics of the turbojet engine MPM-20, 12th IEEE International Symposium on Computational Intelligence and Informatics, 21. - 22. November, Budapest, Hungary, ISBN 978-1-4577-0043-9, pp. 259 – 264, (2011).

Intrusion Detection Systems – Overview, Timeline and Techniques

Marek ČAJKOVSKÝ (1st year), Ivan KLIMEK (2nd year), Michal Ennert (1st year)
Supervisor: Martin TOMÁŠEK

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

marek.cajkovsky@tuke.sk, ivan.klimek@tuke.sk, michal.ennert@tuke.sk, martin.tomasek@tuke.sk

Abstract — Recent high profile network/computer security breaches by various hacker groups (e.g. Anonymous, LulzSec) demonstrated that comprehensively securing such systems is a non-trivial task. The main reason causing this state is probably sheer scale of the topic, lack of an unified security solution which can result in security gaps due to fragmentation of approaches. One of the most promising fields in computer security that could change the current sub-optimal state are intrusion detection (prevention) systems (IDS - IDPS). Their primary purpose is to monitor target computer or networks system and discover potentially dangerous behavior and in case of IDPS directly take appropriate counter measures. This paper aims to be a short overview of various IDS and techniques that they are based on.

Keywords— Intrusion Detection (prevention) System (IDS - IDPS), malicious behavior, computer security, malware

I. SECURITY - INTRODUCTION

According to P. Ning and S. Jajodia [18] intrusions and intrusion detection are defined as: "The activities that violate the security policy of the system, and intrusion detection is the process used to identify intrusions". Another definition says: "Intrusion detection systems monitor system and network resources to detect unusual activity or changes." [16] Vulnerability of computer system can be divided according several factors. Intrusions can originate from various sources, most generally they can be divided into [21]:

- Outsider attacks
- Insider attacks

Outsider attacks are performed by non-authorized users, on the other hand insider attacks are originate from authorized users who abuse their privileges.

II. IDS - OVERVIEW

IDS can be divided according multiple factors which vary from one literature to another. Two main divisions of IDS according to their deployment are [1]:

- Host-based IDS (aka host-oriented IDS)
- Network-based IDS (aka network-oriented IDS)

Host-based IDS

A host-based IDS analyze Operating System (OS) events such as process identifiers and system calls, [1] also host-based IDS track changes made to important files and directories. [16] The biggest challenge is to choose features that best characterize the user or system usage patterns so that non-dangerous activities would not be classified as anomalous. One could use, for instance, Unix Shell command lines, login events or system calls as observables to generate profiles of user behavior. [4]

Network-based IDS

Network-based IDS are placed on the network near the system or systems being monitored and analyze network traffic for attack patterns and suspicious behavior. [16] Interesting project is e.g. open-source Bro IDS - Time machine, this approach allows to record the entire contents of a high-volume network traffic stream, in order to later "travel back in time" and inspect activity that has only become interesting in retrospect. [9] Several network-based IDS such as EMERALD, NFR, BRO and further references are mentioned in [36]

According to Vigna and Kemmerer [22] network-based IDS can be further divided into:

- distributed IDSs
- network-based IDSs

Finding unknown intrusions in network traffic can be very complicated, whereas on a host there are more things to be looked at such as processes, network accesses, system calls etc. [7] IDS can be also divided according to their functionality. [10][37] From this perspective we divide IDS to two main categories:

- Rule-based IDS (aka Signature-based IDS, Knowledge-based IDS or Misuse-based IDS)
- Behavioral-based IDS (aka Statistical-anomaly-based IDS, Statistical-based IDS or Anomaly-Based IDS)

Some literature also mentions different classifications, e.g. Sekar et al. [3] added the specification-based category, Jha and Hassan [5] the agent-based. According to Thamizharasan

et al. [8] the IDS performance in detecting the intrusions more accurately can be improved by using the agent technology with the combination of State Transition Analysis Tool. [21] An agent is able to learn from its experiences. It is autonomous and takes actions based on its built-in knowledge and its past experiences [43] Verma et al. [45] says that an agent can be defined as a software program capable of executing a complex task on behalf of a user.

Uppuluri and R. Sekar [43] showed that effective specification-based IDS can be developed with modest efforts, they described specification-based IDS as a system where manually developed program behavioral specifications (rather than previously seen) are used as a basis to detect attacks. They also showed that specification-based techniques are similar to anomaly detection in that they also detect attacks as deviations from a norm, and thus are able to detect novel attacks. [3]

Rule-based IDS

Rule-based IDS maintain database of patterns (signatures) which are then matched to packets (or some other information unit). Therefore such IDS are only as strong as their database of signatures. [12][13][14] Signature-based detection can be tricked by polymorphism, obfuscation and packing techniques [47] [48], this leads to fact that a new signature is required for almost every single malware variant.

Advantages: potential to generate low false alarm rates

Drawbacks: unable to detect new and previously unidentified attacks

Behavioral-based IDS

Behavioral-based IDS rely on statistical models to identify anomalous behavior. To identify an anomaly, the system uses data compiled from previous network or computer system behavior. This property enables them to learn system/user behavior and detected its deviations. [1][12][13][14][5]

Advantages: capable of detect new and previously unidentified attacks

Drawbacks: potential to generate high false alarm rates, can be mistaken during learning phase

Further categorizing of IDS can be according to information that is gathered from the observed system. This is mostly true for host-based IDS because network-based IDS primary capture network traffic as it was mentioned before. [16]

IDS have been integrated widely with lots of new technologies in many areas, such as data mining algorithm, fuzzy logic theory, protocol analyzing, etc. [8] This may suggest that IDS can be divided also according to method which is used for “finding anomalies” this is primary true for behavioral-based IDS. Garcia-Teodoro et al. [1] explain differences between techniques which are used for anomaly detection. Several comparison studies exist on this topic e.g. Javitz and Valdes [6] discuss various techniques and compare them to their proposed system NIDES.

According to Garcia-Teodoro et al. [1] the main differences between signature-based and anomaly-based IDS are inherent in the concepts of “attack” and “anomaly”. An attack can be defined as “a sequence of operations that puts the security of a system at risk”. Garcia-Teodoro et al. [1] further classified the anomaly detection techniques in behavioral-based IDS according to the nature of the processing to the following

categories and sub-categories, further information to each category can be found in [1]:

- Statistical based
 - Unvariante
 - Multivariante
 - Time series model
- Knowledge based
 - FSM
 - Description language
 - Expert systems
- Machine learning based
 - Bayessian networks
 - Markov models
 - Neural networks
 - Fuzzy logic
 - Genetic algorithms
 - Clustering & outlier detection

As it is mentioned in Okazaki's et al. work [35], Warrender [41] wrote a survey on intrusion detection method based on system calls. He counts four techniques:

1. analyzing a sequence of system calls
2. counting the number of system calls
3. using data mining
4. tracing the state by a finite state machines

Aforementioned Okazaki's et al. work [35] is classified as a 1st technique.

There are also many variations of the mentioned techniques which combine different approaches and overlap, e.g. in the approach proposed by Sekar et al. that combines the specification-based and anomaly-based intrusion detection. [3] Every aforementioned solution has its own advantages and drawbacks. No intrusion detection approach stands alone as a catch-all for computer penetrations; each approach is technically suited to identify a subset of the security violations to which a computer system is subjected. [21] Behavioral-based intrusion detection needs a larger amount of system resources like CPU time, memory and disk space than rule-based intrusion detection. [35] Among the methods adopted by the intrusion detection system, the most significant method is State Transition Analysis Tool (STAT) which can put forward the IDS in an efficient way [42].

III. IDS - TIME-LINE

The following text describes the IDS development time-line and its milestones.

One of the first computer security concepts was published by James P. Anderson in the early 80's [11]. This concept was expanded by D. Denning in his intrusion detection model in 1987 [17]. Both researchers relied on audit trail data that were gathered from the observed system and also considered standard deviation in their models. Proposed audit trail data sources were e.g. frequency of logins or system resources (CPU time I/O operations, etc).

Based on Denning's work the NIDX model was proposed by Bauer and Koblenz [19]. Bauer and Koblenz extended Denning's IDES to include target system depended knowledge. [19]

In 1994 Kumar and Spafford in their work: A pattern matching model for misuse intrusion detection, described generic model of matching that can be usefully applied to misuse intrusion detection. Their model was based on Colored Petri Nets (CPN). [20]

In 1995 Ilgun, Kemmerer and Porras implemented STAT (state transition analysis tool) for Unix like operating systems called USTAT. [21] Both Kumar et al. and Ilgun et al. models system behavior in similar way, they've used fact that the penetration can be modeled as a series of state changes that lead from an initial secure state to a target compromised state. [20]

In 1998 Vigna and Kemmerer presented netSTAT. Their approach was using a formal model of both the network and the attacks. By using this formal model netSTAT was able to determine which network events have to be monitored and where they can be monitored. [22]

In 1998 Hofmeyr et al. in the paper: Intrusion detection using sequences of system calls, strongly affected future research. Hofmeyr et al. introduced detecting intrusion at the level of privileged processes. Their work proved that short sequences of system calls executed by running processes are a good discriminator between normal and abnormal operating characteristics of several common UNIX programs. [23] Further work by Hofmeyr et al. can be found in [24][25].

In 1998-1999 Martin Roesch presented SNORT - libpcap-based packet sniffer and logger that can be used as a lightweight network intrusion detection system. With millions of downloads and nearly 400,000 registered users, Snort has become the de-facto standard for IPS. Snort is a typical example of a rules-based traffic collection engine. [26][27]

In 2001 Wagner and Dean showed how static code analysis may be combined with system call tracing and used to automatically derive a model of application behavior. They solution presumes no logic errors such as array boundary or dereferencing NULL pointer to minimize false alarms. Their result is a host-based intrusion detection system with three advantages: [28]

- high degree of automation
- protection against a broad class of attacks based on corrupted code
- elimination of false alarms

In 2005 Tan and Sherwood developed a string matching approach for intrusion detection that could be included on a chip. This approach was 10 times more efficient than other approaches at the time. [29]

In 2008 Smith, Estan, and Jha introduced extended finite state automata (XFSA) and refer to fact that this approach can be used in IDS. [30]

In 2009 Bayer et al. propose a novel, scalable clustering approach to identify and group malware samples that exhibit similar behavior. They use dynamic binary analysis and generalizing into behavioral profiles. The presented clustering performance was demonstrated on set of more than 75 thousand samples that were processed in less than three hours. [31]

IV. RECENT WORK IN MALWARE CLUSTERING

Observing system calls seems good idea not only on Unix-like operating systems, Bose et al. [46] presented monitoring

system events and API calls on Symbian OS. They work propose an efficient representation of malware behavior based on a key observation that the logical ordering of an application's actions over time often reveals the malicious intent even when each action alone may appear harmless. They used to discriminate malicious behavior from normal behavior by training a classifier based on Support Vector Machines (SVM). They proved that their solution can identify current mobile viruses and worms with more than 96% accuracy. Gurrutxaga et al. [49] proposed malware comparison in an unsupervised learning environment, they also compared several malware clustering algorithms. Kinable and Kostakis [50] presented solution based on call graph clustering, they also compared various clustering algorithms, such as k-medoids and DBSCA, they used set of malware samples from F-Secure Corporation. Perdisci et al. [51] presented an approach that can automate process of extracting network signatures for detecting HTTP traffic generated by malware-compromised machines. Li et al. [52] compared "classical" methods used in malware clustering against methods used in detecting plagiarism. [53] Wang and Jiang [53] were inspired by our natural immune systems that is very effective in protecting our body from the intrusion by (almost endless) variations of pathogens, their work was based on S. Forrest et al. [25] [54] As such, dynamic analysis of malware is often far more effective than static analysis. Monitoring the behavior of the binary during its execution enables collecting a profile of the operations that the binary performs and offers potentially greater insight into the code itself if obfuscation is removed (e.g., the binary is unpacked) in the course of running it. [52] While this technique has its limitations—e.g., it may be difficult to induce certain behaviors of the malware, some of which may require certain environmental conditions to occur. [52] For this reason, dynamic analysis of malware has received much attention in the research community. Analysis systems such as CWSandbox [56], Anubis [57], BitBlaze [55], Norman and Threat-Expert execute malware samples within an instrumented environment and monitor their behavior for analysis and development of defense mechanisms. [52] Caballero et al. proposed a solution that is capable to extract format of protocol messages sent by an application implementing an unknown protocol specification. The proposed technique was successfully used against MegaD, spam botnet without a documented C&C protocol [58] Wang et al. [59] presented ReFormat - system that aims at deriving the message format even when the message is encrypted. They assume that messages go through several phases including message decryption so by monitoring what the system does such as mathematical operations etc. it is possible to decrypt the communication.

V. IDS - PERFORMANCE TESTING

The previous text covers only a subset of all available approaches. As it can be seen the methods greatly differ from each other, and often involve large variety of both computer-science (CS) and non-CS techniques.

According to [32][33] it is hard to tell which technique is the best due to lack of appropriate metrics. One of the possible "workaround" how to test intrusion detection performance and accuracy is to use one of the available measured historical data. According to Zhang et al. [40] there are two datasets that are widely used to evaluate the performance of the intrusion

detection. One is the dataset which was generated by Forrest's research group at New Mexico University and the other one by a research group at MIT. The MIT dataset was used at KDD Cup 1998 and 1999.

For example Okazaki et al. [35], Li et al. [39] and Zhang et al. [40] used data measured on University of New Mexico (available at <http://www.cs.unm.edu/~immsec/data-sets.htm>). Lee, Stolfo [36], Latifur et al. [37] and Sekar et al. [3] used the second mentioned data from the: Cyber Systems and Technology Group (formerly the DARPA Intrusion [Detection Evaluation](#) Group) of MIT Lincoln Laboratory (available at <http://www.ll.mit.edu/mission/communications/CST/darpa.htm>), Uppuluri and Sekar in [44] used both datasets. Further information about DARPA data set can be found at Lippmann et al. [38]

It is important to mention that nearly all of the projects to design or construct a secure systems for processing classified information have had a formal mathematical model for security as part of the top-level definition of the system. [34]

REFERENCES

- [1] Garcia-Teodoro, P. and Diaz-Verdejo, J. and Macia-Fernandez, G. and Vazquez, E. - 2009 - Anomaly-based network intrusion detection: Techniques, systems and challenges
- [2] Debar, H., Dacier, M., Wespi, A. - 1999 - Towards a taxonomy of intrusion-detection systems
- [3] Sekar et al. - 2002 - Specification-based anomaly detection a new approach for detecting network intrusions
- [4] Liao, Vemuri - 2002 - Use of K-Nearest Neighbor classifier for intrusion detection1
- [5] S. Jha, M. Hassan - 2001 - Building agents for rule-based intrusion detection system
- [6] Javitz, Valdes - 1993 - The NIDES statistical component Description and justification
- [7] Richardson, Trajtkovski - 2007 - An Agent-Based Intrusion Detection System
- [8] Thamizharasan et al. - 2011 - A Feasible Study On Agent-Based Intrusion Detection System
- [9] BRO IDS: TIME MACHINE - 12/03/09 - official site - <http://tracker.bro-ids.org/time-machine/>
- [10] HINES, Michael R. - Going Beyond Behavior-Based Intrusion Detection
- [11] Anderson - 1980 - Computer Security Threat Monitoring and Surveillance, Fort Washington, PA (Apr. 1980)
- [12] Jamil Farshchi - 12/03/12 - Intrusion Detection FAQ: Statistical based approach to Intrusion Detection - http://www.sans.org/security-resources/faq/statistic_ids.php
- [13] Andy Cuff - 12/03/12 - Intrusion Detection Terminology (Part One) - <http://www.symantec.com/connect/articles/intrusion-detection-terminology-part-one>
- [14] Andy Cuff - 12/03/12 - Intrusion Detection Terminology (Part Two) - <http://www.symantec.com/connect/articles/intrusion-detection-terminology-part-two>
- [15] Jamil Farshchi - 12/03/12 - Statistical-Based Intrusion Detection - <http://www.symantec.com/connect/articles/statistical-based-intrusion-detection>
- [16] Allison Hrivnak: Host Based Intrusion Detection: An Overview of Tripwire and Intruder Alert, January 29, 200, SANS Institute InfoSec Reading Room
- [17] Denning - 1987 - An Intrusion-Detection Model
- [18] Ning, Jajodia - 2003 - Intrusion detection techniques. The Internet Encyclopedia
- [19] Bauer, Koblenz - 1988 - NIDEX-an expert system for real-time network intrusion detection
- [20] Kumar, Spafford - 1994 - A pattern matching model for misuse intrusion detection
- [21] Ilgun, Kemmerer, Porras - 1995 - State transition analysis A rule-based intrusion detection approach
- [22] Vigna, Kemmerer - 1998 - NetSTAT A network-based intrusion detection approach
- [23] Hofmeyr, Forrest, Somayaji - 1998 - Intrusion detection using sequences of system calls
- [24] Somayaji, Hofmeyr - 1998 - Principles of a computer immune system
- [25] S. Forrest and S. A. Hofmeyr and T. A. Longstaff and A. Somayaji - A Sense of Self for Unix Processes
- [26] SNORT IDS - 12/03/09 - official website - <http://www.snort.org>
- [27] Roesch - 1999 - Snort-lightweight intrusion detection for networks
- [28] Wagner, Dean - 2001 - Intrusion detection via static analysis
- [29] Tan, Sherwood - 2005 - A high throughput string matching architecture for intrusion detection and prevention
- [30] Smith, Estan, Jha - 2008 - XFA Faster signature matching with extended automata
- [31] Bayer et al. - 2009 - Scalable, behavior-based malware clustering
- [32] Stolfo, Fan - 2000 - Cost-based modeling for fraud and intrusion detection: results from the JAM project:130.44.
- [33] Gaffney, Ulvila 2001 - Evaluation of intrusion detectors: a decision theory approach:50.61.
- [34] Landwehr - 1981 - Formal models for computer security
- [35] Okazaki, Sato, Goto - 2002 - A new intrusion detection method based on process profiling
- [36] Lee, Stolfo - 2000 - A framework for constructing features and models for intrusion detection systems
- [37] Latifur Khan, Mamoun Awad, Bhavani Thuraisingham - 2007 - A new intrusion detection system using support vector machines and hierarchical clustering
- [38] Richard Lippmann, Joshua W Haines, David J Fried, Jonathan Korba, Kumar Das - 2000 - The 1999 DARPA off-line intrusion detection evaluation
- [39] Jimin Li, Zhen Li, Kunlun Li - 2009 - Computer system security model based on system call related to security
- [40] Zhang, Zhu, Fan - 2005 - Intrusion detection based on cross-correlation of system call sequences
- [41] Warrender, Forrest, Pearlmuter - 1999 - Detecting intrusions using system calls Alternative data models
- [42] Bin Dong, Xiu-Ling Liu - 2007 - An Improved Intrusion Detection System Based on Agent
- [43] I. M. Hegazy, T. Al-Arif, Z. T. Fayed, H. M. Faheem - 2003 - A multi-agent based system for intrusion detection, Potentials, IEEE, vol. 22, no. 4, pp. 28–31, 2003
- [44] P. Uppuluri and R. Sekar - 2001 - Experiences with Specification-based Intrusion Detection
- [45] Verma, Husain, Shukla - 2011 - Research on Mobile agent based network intrusion
- [46] Bose et al. - 2008 - Behavioral detection of malware on mobile handsets
- [47] M.Christodorescu, S.Jha, S.A.Seshia, D.Song, and R.E.Bryant. Semantics-aware malware detection. In Proceedings of the IEEE Symposium on Security and Privacy, 2005.
- [48] J.A.Morales,P.J.Clarke,Y.Deng,andB.M.G.Kibria. Testing and evaluating virus detectors for handheld devices. Journal in Computer Virology, 2(2):135–147, November 2006.
- [49] Gurrutxaga et al. - 2008 - Evaluation of malware clustering based on its dynamic behavior
- [50] Kinable, Kostakis - 2010 - Malware Classification based on Call Graph Clustering
- [51] Perdisci, Lee, Feamster - 2010 - Behavioral clustering of HTTP-based malware and signature generation using malicious network traces
- [52] Li et al. - 2011 - On challenges in evaluating malware clustering
- [53] Wang, Jiang - 2011 - Artificial malware immunization based on dynamically assigned sense of self
- [54] S. Forrest, S. Hofmeyr, and A. Somayaji. Computer Immunology. Communications of the ACM, 40(10):88–96, 1997
- [55] D. Song, D. Brumley, H. Yin, J. Caballero, I. Jager, M. G. Kang, Z. Liang, J. Newsome, P. Poosankam, and P. Saxena. Bitblaze: A new approach to computer security via binary analysis. In Proceedings of the 4th International Conference on Information Systems Security, December 2008.
- [56] C. Willems, T. Holz, and F. Freiling. Toward automated dynamic malware analysis using cwsandbox. In Proceedings of the 2007 IEEE Symposium on Security and Privacy (S&P07), pages 32 – 39, 2007.
- [57] U. Bayer, C. Kruegel, and E. Kirda. Ttanalyze: A tool for analyzing malware. In 15th European Institute for Computer Antivirus Research (EICAR 2006) Annual Conference, 2006.
- [58] Caballero, J., Berkeley, U. C., Poosankam, P., Song, D., & Berkeley, U. C. (2009). Dispatcher : Enabling Active Botnet Infiltration using Automatic Protocol Reverse-Engineering. Security.
- [59] Wang, Z., Jiang, X., Cui, W., & Wang, X. (2009). ReFormat: Automatic reverse engineering of encrypted messages. North. Retrieved from <http://www.springerlink.com/index/e12g64j4855u1106.pdf>

Jitter Utilizing Congestion Estimation (JUICE)

Ivan KLIMEK (2nd year), Marián Keltika (1st year external)
 Supervisor: Assoc. Prof. František Jakab, PhD.

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

ivan.klimek@tuke.sk, keltika@keltika.sk, frantisek.jakab@tuke.sk

Abstract—This paper presents a novel approach to congestion control for unicast transport protocols utilizing jitter in combination with packet drop to detect and utilize the flow synchronization properties of the most commonly used FIFO/DropTail queue scheduling/drop policies. The first simulationally obtained findings contradict the standard definition of TCP-Friendliness which dictates that all flows have to react to packet drop as TCP does to guarantee fair bandwidth sharing. Drop/Jitter anomalies are studied and first results presented, proving that it may be possible to provide more stable bandwidth allocation but at the same time be backward compatible with existing TCP variants.

Keywords—congestion control, jitter, queue scheduling, flow synchronization

I. INTRODUCTION

Recent advances in rateless erasure codes [1, 2] enable to separate reliability from rate/congestion control for general purpose unicast transport protocols [3]. Currently for purposes where reliability is not necessary (for example multimedia applications, or when other layers take care of reliability for example through AL-FEC) but congestion control has to be guaranteed - TCP-Friendly Rate Control (TFRC) provides an option [4]. But it does so by modeling the TCP performance as a function of packet loss and RTT under the same conditions and limiting its rate accordingly. We argue that this approach is sub-optimal because it inherits the TCP performance which is not always satisfactory¹. As an alternative we propose to use jitter as an additional control signal in combination with the currently used packet drop. We explore the behavior of jitter before and after congestion events on the most widely used FIFO/DropTail queue scheduling/drop policies [5] and propose a mechanism to detect flow synchronization and use it to discover the dynamic capacity equilibrium where maximal rate is achieved without being too aggressive and limit other flows. The proposed mechanism is verified using ns-2 simulations.

II. JITTER AS CONTROL SIGNAL

Several variants of TCP already proposed using RTT as an additional control signal on top of packet drop.

TCP Vegas [6] aims to improve the end-to-end congestion avoidance mechanism of TCP. The main objective is to estimate the expected bandwidth for the connection in order to control the transmission rate that can avoid network congestion. To achieve this goal, the scheme defines *BaseRTT* value which represents the minimal round trip time

¹For example on Wireless WAN links, or Long Fat Networks (LFN), resp. the issue of Bufferbloat etc.

during the transmission to calculate the expected transmission rate of the link. After receiving an acknowledgement, sender

continues to update *ActualRTT* value which represents the current RTT and is necessary to be able to calculate the real transmission rate. If the difference between *BaseRTT* and *ActualRTT* is higher than upper bound threshold, congestion may occur since sending rate is too high. Thus, sender decreases one congestion window size. If the difference is smaller than lower bound, sender should increase one congestion window size to utilize the available bandwidth. Else, sender should keep the sending rate stable. TCP Vegas suffers fairness problems when the connections start transmitting at different times. Also, TCP Vegas is not suitable for wireless network since it cannot distinct loss events.

TCP Veno [7] is a loss differentiation scheme for the wireless environment and it is derived from TCP Vegas. This method provides another threshold to differentiate between wireless and congestion losses. Although the performance is improved in wireless environment, the loss differentiation scheme cannot work well when the random loss rate is high. And it still does not solve the fairness issues of Vegas.

TCP Jersey [8] is another enhancement which improves network performance in wireless network. But because it needs router support (ECN) we will not describe it further.

All the mentioned schemes so far used RTT, such congestion estimator cannot perform well when the traffic load gets heavy over reverse links or asymmetric links.

On the other hand, JTCP [9] applies the jitter ratio to differentiate wireless losses from congestion losses and revise the Reno's congestion control scheme to adapt to wireless environments. Jitter ratio [10] is derived from the inter-arrival jitter, which is defined in Real-time Transport Protocol (RTP)

[11]. Inter-arrival jitter is the variance of packet spacing at the receiver side and packet spacing at the sender side. In other words, it presents current path's status by the packet-by-packet delay. On packet loss JTCP compares the average jitter ratio with a threshold, which is defined as the inverse of congestion window size. If average jitter ratio is greater than this threshold, JTCP regards the loss event as congestion loss and reduces its congestion window size to one half. Otherwise, the loss event will be viewed as wireless loss. Sender will not reduce the congestion window and will do fast retransmission immediately.

Jitter seems to be better suited as an additional control signal than RTT as it describes only the one way delay changes triggered by changed path load, thus it is more stable

as it cannot be affected by the load of the reverse path.

III. JITTER UTILIZING CONGESTION ESTIMATOR (JUICE)

The main idea behind JUICE is that using Inter-arrival jitter it is possible to gain enough information about the link/bottleneck queue, so that the traditional TCP additive increase multiplicative decrease (AIMD) behavior can be modified to be more stable, utilize the available bandwidth faster and still be TCP-Friendly. The primary principle used is the behavior of FIFO/DropTail queues that tend to synchronize flows on buffer overflow, as the majority of flows are TCP-Friendly they will drop their sending rate. This causes the queue to free up and the one way delay thru the loaded link to shrink creating “negative” jitter. Our ns-2 experiments show that such jitter behavior closely following small² amounts of packet drop is a clear indicator of flow synchronization. Instead of slowing down to half of the sending rate, JUICE reduces the sending rate only to the amount of experienced packet loss and holds the sending rate stable giving other flows the opportunity to catch up. If JUICE would increase the rate even after flow synchronization, all the TCP-Friendly flows would starve. Thus, it is critical for JUICE to stop increasing the rate and monitor the jitter. If there are other competing flows increasing their rate, it will be visible on the jitter. If the jitter stays stable for prolonged periods of time³ JUICE will slowly try to increase its rate as stable jitter is understood as a signal that more capacity is available. Only on repeated loss events or drop above a threshold currently set to 2 percent loss per RTT the rate is halved exactly as in TCP.

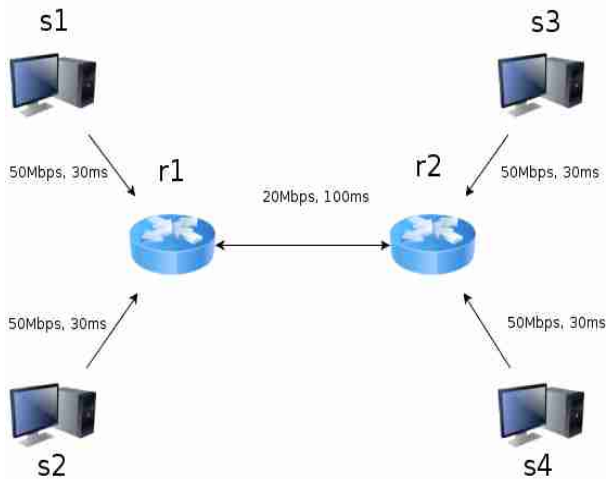


Figure 1: The test ns-2 topology

Figure 1 illustrates the topology used in the ns-2 simulations, node s1 generated the test traffic which was using UDP packets to node s3 acting as the receiver. Node s2 runs a simulated TCP transfer to node s4 acting as the receiver. Bottleneck queue was located at r1 which used the DropTail policy with queue limit set at 200 packets.

Figure 2 demonstrates the performance of JUICE vs. TCP RENO both streams starting and ending at the same time. Figure 3 demonstrates the same situation, only the TCP stream starts at time 20 seconds and ends at time 100

seconds. Figure 4 shows 2 TCP stream vs. one JUICE stream. TCP RENO was chosen because it is the most widely deployed TCP variant. [7]

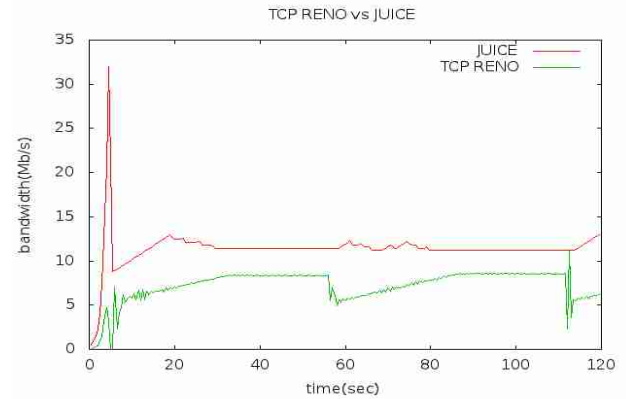


Figure 2: TCP RENO vs. JUICE

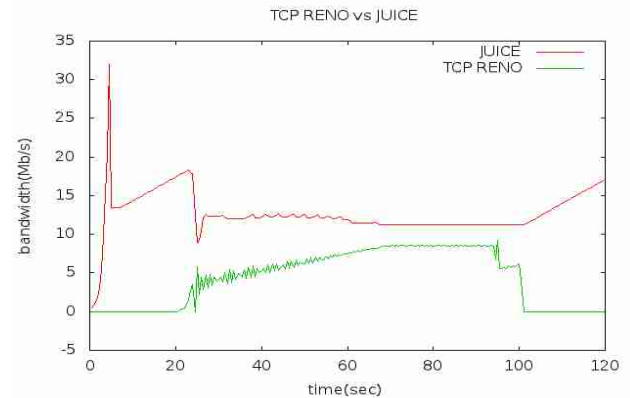


Figure 3: TCP RENO different start/end vs.

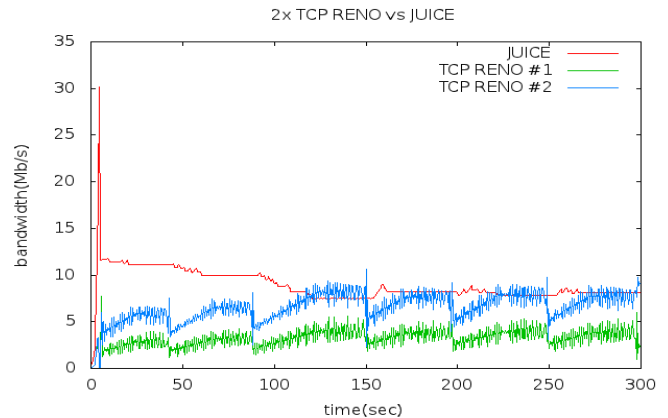


Figure 4: 2x TCP RENO vs. JUICE

IV. CONCLUSION

The first simulation results of the proposed Jitter Utilizing Congestion Estimator (JUICE) show promise into further research even if currently not complete TCP-Fairness can be achieved - JUICE is still on average about 20 percent more aggressive than TCP RENO. The performance on simulated links with non-congestion caused losses remains yet to be simulated, but because of the fundamental principle of loss/jitter congestion estimation behind JUICE we expect more stable behavior when compared to TCP. An experimental real-world testing platform is also under development.

²A more precise definition has yet to be formulated.

³For testing purposes 5x RTT was used.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] M. Luby, A. Shokrollahi, M. Watson, T. Stockhammer, L. Minder. "RaptorQ Forward Error Correction Scheme for Object Delivery." RFC 6330. August 2011.
- [2] A. Shokrollahi, M. Luby. "Raptor Codes (Foundations and Trends(R) in Communications and Information Theory)." Now Publishers Inc., May 2011. 978-1601984463, 10.1561/0100000006.
- [3] I. Klimek. *Wide Area Network Traffic Optimization*. Written work to Doctoral Thesis. Technical University Košice, December 2011.
- [4] S. Floyd, E. Kohler, J. Padhye. "Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 3: TCP-Friendly Rate Control (TFRC)." RFC 4342. March 2006.
- [5] Rade Stanojević, Robert Shorten, "Trading link utilization for queueing delays: An adaptive approach," *Computer Communications*, Volume 33, Issue 9, 1 June 2010, Pages 1108-1121, ISSN 0140-3664, 10.1016/j.comcom.2010.02.014.
- [6] L. S. Brakmo and L. L. Peterson, "TCP Vegas: end to end congestion avoidance on a global internet," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 8, pp. 1465–1480, 1995.
- [7] C. P. Fu and S. C. Liew, "TCP Veno: TCP enhancement for transmission over wireless access networks," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 216–228, 2003.
- [8] K. Xu, Y. Tian, and N. Ansari, "TCP-Jersey for wireless IP communications," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 4, pp. 747–756, 2004.
- [9] E. H. K. Wu and M. Z. Chen, "JTCP: Jitter-based TCP for heterogeneous wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 4, pp. 757–766, 2004.
- [10] S. Y. Chen, E. H. K. Wu, and M. Z. Chen, "A new approach using time-based model for TCP friendliness rate estimation," in *Proceedings of the International Conference on Communications (ICC '03)*, pp. 679–683, May 2003.
- [11] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Application," Internet Engineering Task Force (IETF), RFC 1889, 1996.

Management of enterprise security

¹Ján ŠTOFA (1st year), ²Dana PALOVÁ

Supervisor: ³Kristína MACHOVÁ

^{1,3}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

²Dept. of Economics, EKF TU of Košice, Slovak Republic

¹jan.stofa@tuke.sk, ²dana.palova@tuke.sk, ³kristina.machova@tuke.sk

Abstract — the article focuses on Enterprise Security and emphasizes the importance of research which has been undertaken by FOCUS and Deloitte. It examines policy, the role of management, and the various threats to security. Recommendations at company and legislative level are presented. The main role is to maintain and manage existing security by using every resource available. This would entail examination of current strategies and the identification of improved methods to sustain effective management of security in the future.

Keywords — management, security, enterprise, security project, process, threat

I. INTRODUCTION

According to the survey realized by FOCUS agency, the most dynamically growth area in the world is focusing on security. The results of the survey show the most significant shift occurred in the awareness of the importance of this huge problem. With the current development of IT followed by the expansion of possible threats becomes to the forefront the effort of enterprises to design integrated safety philosophy. FOCUS agency carried out similar surveys focused on enterprise security issue during the years 2004, 2006 and 2008. Comparison of achieved results of particular surveys proved the progress in an effort to increase business efficiency of providing safety in its' environment, but at the same time showed list of various new threats needed to be eliminated [1], [3].

At the international level was similar survey focused on security in the technology, media and telecommunications realized by Deloitte Company in 2011. The main conclusion of this survey is that the threat of information technology increases and the impact of security incidents are still significant. It was also stated that the greatest threat are mobile technologies (e.g. mobile phones and tablets). In particular, the most significant is possibility of loss the sensitive data that are usually stored in these apparatus [4].

Security as a condition for operation of the company is subject to control and competence of executives and other responsible persons. In terms of providing sustainable business prosperity, it is necessary to introduce to enterprise environment the steering system - management strategy [5]. One of the parts of management strategy is presented by security management. The main role is to maintain the security and continuously and effectively use all of resources providing stable operation of the enterprise at present and in the future. This requires a proactive approach of managers to be able to:

- identify immediate threat,
- identify possible threats of security,
- detect possible threats of security,
- develop security system available effectively eliminate or minimize potential risk of security incidents of all assets of the company.

II. ENTERPRISE SECURITY ASPECTS

The safety of the enterprise can be viewed from different angles of views, e.g. by categories:

- Physical, objective and natural category - fire, earthquake, storm, land subsidence, lightning, etc.,
- Personnel category (the so-called human factor) - the violent intrusion into an object, intentionally damage by employees, users error, theft, and staff shortages etc.,
- Communication category - listening, wrong targeting of the messages, etc,
- Technical information category - definition of user rights (using software illegally, downloading and installing unauthorized software), data destruction, archiving and backup of data, unauthorized access to network, defect conduction of electrical current, etc.

The above-mentioned survey by Deloitte, among other things, define the five major threats to IT security (Fig.1):

- Mobile devices (31%),
- Security violations involving third part (23%),
- Mistakes and negligence of employees (17%),
- Rapid implementation of emerging technologies (15%),
- Abuse of IT systems and information from the part of employees (14%) [1], [7].

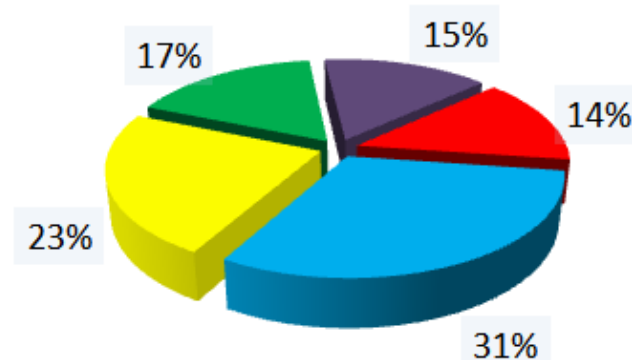


Fig.1 The proportion of different types of threats. Source: [1]

An important aspect of countering the threats in the enterprise is their own employees. They often do not reflect how can endanger business or their work. For example, they are often unreasonably "surfing" on the Internet, which may allow the virus infiltrations. Those depend on the type of affect and functionality of the system or degrade or exploit a variety of protected data. The most commonly used techniques constituting threats and possible attacks on software and network resources are:

- IP spoofing,
- Packet sniffing,
- Highjacking,
- Social engineering,
- Buffer overflow,
- Spamming,
- Password cracking,
- Trojan,
- Worm,
- Logic bomb,
- Malware [8].

The following schema (Fig 2) shows the factors affecting the level of risk and threat which must be resisted. In practice, the management tests all possible risks to its external and internal structures to reduce them to an acceptable level by improvement and recommendation.

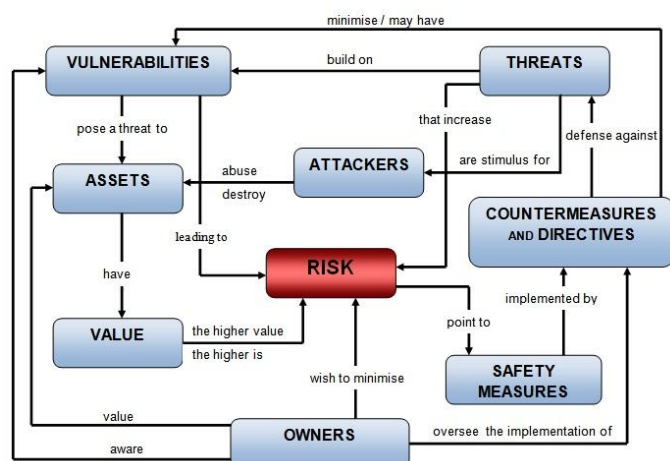


Fig.2 Description of system of functions of the safety risk factors,
Source: [9]

Risk in general presents a correlation between the probabilities of creation a negative phenomenon and its consequences. Mathematically, this relationship can be expressed as follows:

$$R = P \times D \quad (1)$$

where:

R – degree of risk

P – probability of the event,

D – consequence of the observed events [10].

The enterprise has to define clearly all ongoing processes. This is the way how should enterprise operate as efficiently as possible with full use of all its resources. Business processes and their precise and clear definition and description significantly affect the safety of the personal data and sensitive information. Business processes influences affect in particular

way to administration and decision making process [11].

The quality of the processes in the system of quality management plays a significant role [12], [13]. For the enterprise it becomes objective that directs all of the processes/actions in the enterprise. It's the way how the enterprise ensures smooth running and its competitiveness in the sector on the market place. In particular, it ensures the requirements for the protection of personal data, sensitive information and classified information.

In determining the hierarchy of their business processes the enterprise can utilize a map of processes as a functional and effective tool (Fig. 3). Business processes can be divided into three basic components:

1. management processes,
2. main processes,
3. auxiliary (supporting) processes.

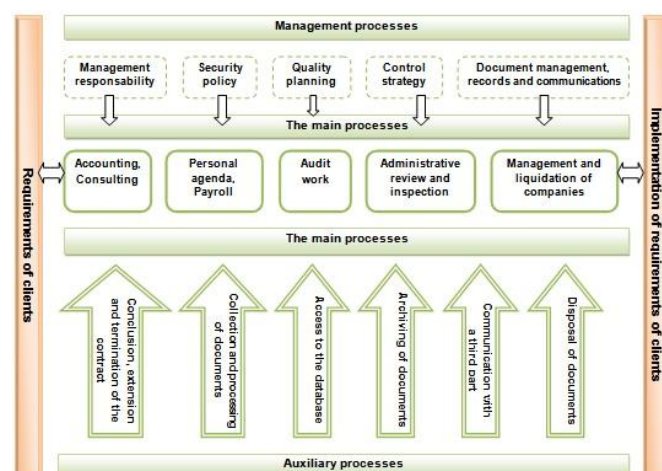


Fig.3 Map of business processes of analyzed enterprise, Source: author

Every enterprise needs to define the implemented processes and include them to the security policy of their organization [14]. If missing security policy in the enterprise or is inconsistently observed it leads to successful attacks on the external and the internal structure of the enterprise.

This security policy constitutes a clear and flexible document taking into account all aspects of the operation of the enterprise and its protection. The Security project serves as an internal law that clearly defines the security of the enterprise. It contains the main plan, its detailed description, as well as other significant detail plans, descriptions, instructions and procedures. The development of the enterprise security policy document is an important prerequisite for the development of security and at the same time the condition of the project [15]. It is largely structured into categories, according to the area, which the security immediately touches:

- personnel security,
- information security,
- physical and object security,
- security of technical means,
- encryption and protection of the information,
- administrative security.

In general, the Security project is a written document in which are specified means, methods and procedures used or recommended for the protection of classified information with regard to specific business conditions. The solver bases on the designated security policy. Fig. 4 offers an insight into the internal structure of the Safety project.



Fig.4 Internal structure of the security project, Source: [16]

The internal structure of the Safety project can be considered as complicated relational formula. The Security project is a kind of sum of input safety analysis which cooperates with internal security policy document. The following three seemingly separate parts (security plan, security analysis, security directives) create the whole of the Security project, which itself is a mandatory obligation derived from legislation.

III. CASE STUDY

Security management we applied to the specific enterprise doing business within the EU in the areas of accounting as an internal part of the security policy. The management of the safety follows and is based on Security project for a particular enterprise. The formation of this document has been preceded by a detailed analysis of the entire enterprise. It starts from external influences affecting the enterprise from the outside, to his innermost details regarding the top classified information.

Thanks to the in-depth analysis of the automated and non - automated information system we managed to find the critical point, which may suggest a threat to the enterprise and its customers and becomes the target of possible attacks. It is analysis and a clear overview of the assets and the sub assets, which the company has.

On the basis of the following analyses we have created a draft of the measures, which contains the steps for correct handling of personal information, sensitive information, and classified information. The proposal of these safety measures was the first step for the development of Security project directives. These Directives point to the possible ways of protection of personal data and accurately detail them: strict adherence to rules of conduct competency and imposed restrictions on employees working on personal data, confidentiality and network restrictions as follows.

Network restrictions:

- do not use wifi connection as far as possible, because the safety of this network cannot be guaranteed; through the wifi connection is possible access to router, where the attacker is able to break the password and become so "lord" of router,
- do not display and change the name of the network (SSID),
- filtering of physical addresses using MAC address,
- use encryption security of communication using WPA2-PSK,
- use encryption at least TKIP (Temporal Key Integrity Protocol Das), but the safety solution is to use AES (Advanced Encryption Standard),
- use password Access - generate a strong password containing at least eight alphanumeric characters and change it at least once a month.

Similar directives – countermeasures, which were proposed to repeal or limit vulnerability to counter threats, were classified into the following three categories:

- protection of physical character,
- protection of a procedural character,
- protection of technical-information character.

Implemented Security project as an internal document for enterprise prepared under current legislation of the State, aims to eliminate, or at least minimize, the risks and threats to the company. In addition, this is of huge benefit for the enterprise models of running business processes and also provides easier training for new workers thus increasing the awareness of the users. It also facilitates the process of reciprocal contexts between processes of corporate audits.

The Security project represents final product of analytical part for the solutions of security of the information system of protection personal data. It summarizes results of the analyses and together with security directives presented means solutions for all levels of security with a description of the security measures. For these reasons, Security project is necessary to be considered a confidential document. Disclosure of its contents by unauthorized persons may result in the elimination of security mechanisms of informational systems and threat to confidentiality.

The following recommendations made as a result of consultations with the leadership of the enterprise and the audit carried out, e.g.:

- installation of video doorman,
- implementation of software into information system to increase the safety of personal information,
- reorganization of the interior equipment,
- increase the safety of personal information,
- strict control of the communications employee – client,
- strict control of the method of obtaining data.

In case of requesting the Security project from the part of Control Commission (Office for the protection of personal data) the enterprise is cover with written confirmation of Control Commission that the document was taken and handed back.

IV. CONCLUSION

The problem of developing security project is very complex and difficult and during its design process it requires huge amount of time. It is impossible to develop it without direct cooperation with analyzed company. Many of security project designers try to generalize this process into global template applicable to all kind of enterprises. We consider that it is big mistake from effective security management point of view. In that case there is present high probability of low level the finally designed Security project. The main reason is the uniqueness of the enterprises, variability of business field, perspective, technical and physical security, and staffing as well as internal and external policies of the enterprise.

One of such forms simplification of the complex analysis of enterprise is utilization of questionnaire forms of analysis. Its provides only incomplete informations which are based on subjective feelings and assumptions of the respondents. Enterprises possess a huge amount of personal data and sensitive information, which is not admissible, in a similar manner, to simplify. It is important to take care of the immediacy, consistency and detail of analysis as the external as well as internal environment and the complex structure of the enterprise.

For an appropriate method of analysis and follow creation of the Safety project we consider integration of „solver“ in the particular work team. His aim is to be understood all ongoing processes and elements of the enterprise concerning the protection of data, persons and property. The Safety project is underpinned by legislation as an instrument in the management of enterprise security.

Currently, there are many enterprises entered to the world of social networks. They allow the handling of information by their openness and easy access. This area is the new phenomenon of the threats and risks of safety as a whole. In an effort to develop the issue of security of personal data and sensitive information, not only in the context of a particular enterprise, we would like to follow up on this issue at a later date. Not only is security of the enterprise an issue but also that of social networking. Significant in this, is the consideration of data mining, information retrieval, opinion analysis, text mining and information filtering, which allows us to track the communication and exchange of information between the actors of the network [17], [18]. The employment of such methods would ensure protection against future threats to both the person and the enterprise.

ACKNOWLEDGMENT

The work presented in this paper was supported by the Slovak Grant Agency of Ministry of Education of the Slovak Republic within the KEGA 065TUKE-4/2011 project (50%). This work is also the result of the project implementation Development of the Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the R&D Operational Program funded by the ERDF (50%).

REFERENCES

- [1] "Firmy sa musia viac zamerať na zvýšenie ochrany dát" [on-line], Available at: <http://www.efocus.sk/kategoria/ict-technologie/clanok/firmy-sa-musia-viac-zamerat-na-zvysenie-ochrany-dat/>, last visit: 2012.3.4, in Slovak
- [2] "Informačná bezpečnosť" [on-line], Available at: <http://www.slideshare.net/jozo/inf-bezpecnost-l-presentation>, last visit: 2012.3.4, in Slovak
- [3] J. Lipianska, J. Hlavatý: "Informačná bezpečnosť podniku v kontexte krízového vývoja hospodárstva" [on-line], Available at: http://of.euba.sk/zbornik2011/ZBORNÍK%20VEDECKÝCH%20STATÍ%202011-PDF/KIOF/LIPIANSKA_J._HLAVATÝ%20C3%9D_I._KIOF.pdf, last visit: 2012.2.20, in Slovak
- [4] "Raising the Bar 2011 TMT Global Security Study – Key Findings" [on-line], Available at: http://www.deloitte.com/assets/Dcom-Croatia/Local%20Assets/Documents/2011/TMT_2011_Global_Security_Survey_hr.pdf, last visit: 2012.2.25, in English
- [5] B. Ulík: "Bezpečnosť ako nikdy nekončiaci proces" In: INFOWARE 10/2009, Bratislava, 2009, s. 18, ISSN: 1335-4787
- [6] R. Gorbý: "Bezpečnostné hrozby malých podnikov" In: ITnews [on-line], Available at: <http://www.itnews.sk/2012-01-31/c146330-bezpecnostne-hrozby-malych-podnikov>, last visit: 2012.3.2, in Slovak
- [7] "GFI Software: v roku 2011 vzrastú bezpečnostné hrozby prostredníctvom výmenných pamäťových médií" [on-line], Available at: <http://www.inet.sk/clanok/11026-gfi-software-v-roku-2011-vzrastu-bezpecnostne-hrozby-prostrednictvom-vymennych-pamatovych-medii/>, last visit: 2012.2.3, in Slovak
- [8] "Škodlivý softvér" [on-line], Available at: <http://www.skodlivysoftware.cz/informace/slovnicek>, last visit: 2012.2.20, in Slovak
- [9] "Common Criteria for Information Technology Security Evaluation" [on-line], Available at: <http://www.commoncriteriaportal.org/files/ccfiles/ccpart1v2.3.pdf>, last visit: 2012.3.2, in English
- [10] "Stupeň rizika" [on-line], Available at: http://www.ebts.besoft.sk/part_UVOD/odborne_forum/.../BP_ARB.doc, last visit: 2012.2.12, in Slovak
- [11] V. Řepa: "Podnikové procesy, Procesní řízení a modelování" Grada Publishing, a.s., Praha, 2007, ISBN 978-80-247-2252-8.
- [12] "Politika kvality a cieľ kvality" [on-line], Available at: <http://manazment-kvality.sk/>, last visit: 2012.2.22, in Slovak
- [13] I. Zolotová, M. Bakoš and L. Landryová: "Possibilities of Communication in Information and Control Systems" In: Annals of the University of Craiova, Vol. 4 (31), no. 2, 2007, 163-168 ISSN 1841-0626.
- [14] I. Zolotová, P. Kubičko, L. Landryová and R. Hošák: "A design of a reference model of an innovation process and its implementation in business using an innovation zone" In: APMS 2011: International Conference on Advances in Production Management Systems: Value Networks: Innovation, Technologies and Management, Stavanger, Norway, 26-28 September, 2011, University of Stavanger, 2011, 1-9, ISBN 978-82-7644-461-2.
- [15] Úrad na ochranu osobných údajov [on-line], Available at: <http://www.dataprotection.gov.sk>, last visit: 2012.3.2, in Slovak
- [16] "Bezpečnosť podniku" [on-line], Available at: vs-bm.snadno.eu/Bezpecnost_podniku.doc, last visit: 2012.2.29, in Slovak
- [17] M. Mach, K. Machová: "Knowledge Technologies for Information Acquisition and Retrieval" Proceedings of the III. ISC'2003 – III. Internal Scientific Conference of the Faculty of Electrical Engineering and Informatics, Košice, 2003, 61-62, ISBN 80-89066-65-8.
- [18] K. Machová, P. Bednár and M. Mach: "Various Approaches to Web Information Processing" Computing and Informatics, Vol. 26, No. 3, 2007, 301-327, ISSN 1335-9150.

Markerless augmented reality using SURF method

Martin VARGA (1st year)
Supervisor: Branislav SOBOTA

Dept. of Computers and Informatics, FEEI TU of Košice, Slovak Republic

martin.varga@tuke.sk, branislav.sobota@tuke.sk

Abstract—This paper presents selected parts of a semi-markerless augmented reality application based on recognition of predefined patterns of square or rectangular shapes in input images. The paper is divided into two main parts. First, augmented reality and its technology are presented. This part also describes difference between marker and marker-less augmented reality. The second part focuses on the pattern recognition procedure using the SURF (Speeded Up Robust Features) method and vertex determination from the input image. These vertexes are necessary inputs of the transformation matrix calculation. Finally, usability of the presented principles is illustrated by an augmented reality application output example.

Keywords—Augmented reality, semi-markerless system, SURF method, pattern detection.

I. INTRODUCTION

There are several technologies suitable for virtual reality (VR) application creation. One of these technologies is mixed reality. Mixed reality (MR) is a computer science research area that points out on the combination of real world with generated data (virtual reality) to achieve high fidelity in VR. The computer generated graphical objects are blended into the real environment in real time. Real world data are provided through sensors in real time back to the mixed reality system [1], [7]. Mixed reality can, according to [2], originate from at least one of the following technologies: augmented reality (AR) and augmented virtuality (AV).

AR environments consist of real world components as well as virtual components. For example, a person using an AR system is equipped with human interface (semitransparent glasses, head mounted display, data display helmet or combination of monitor and camera). Via this interface, the person sees the real world with superimposed computer generated objects on top of the real ones.

Augmented virtuality is technologically similar to augmented reality. However, its principle is opposite to the one of AR. Most of the world (scene) that is displayed is a collection of virtual objects and the real objects are inserted into this world. If the observer is being inserted into the AV scene, he or she is dynamically integrated into the AV system in the same manner as other real objects. That allows real time manipulation with virtual and real objects inside the scene.

Both augmented reality and augmented virtuality systems

are quite similar and they belong under the mixed reality systems' definition. A goal of mixed reality system is to merge real world with the virtual one into a new environment, where real and virtual objects exist together and interact in real time. Relationship between mixed reality, augmented reality and augmented virtuality is shown in Fig.1 [2], [3].

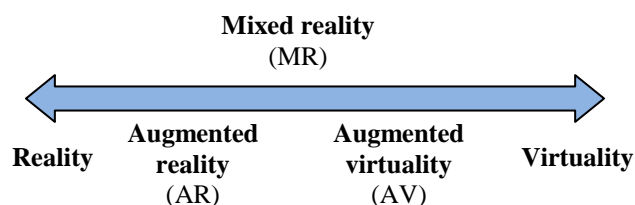


Fig. 1. Milgram's definition of real to virtual world transition

Based on how a user sees mixed reality, there are two types of systems:

- *Optical see-through* systems, where the user sees the real world directly and computer generated objects are added to this view. This category of systems usually works with semi-transparent displays.
- *Video see-through* systems, where the real world is captured by a camera, the image is extended by additional virtual objects and displayed to the user. This is usually realized via camera – display systems [4].

According to the method how virtual objects are aligned with the real scene image, there are two types of systems:

- *Marker systems* – special markers are used in the real scene. The markers are then recognized during runtime and replaced by virtual objects.
- *Markerless systems* – processing and inserting of virtual objects is done without special markers. Additional information is needed, for example a second image (e.g. photo (semi-markerless system)), face recognition, GPS data, etc.

II. SEMI-MARKERLESS AR SYSTEM

A semi-markerless augmented reality application was developed for predefined pattern determination from input images. The architecture of the semi-markerless augmented reality system is shown in Fig. 2. This architecture consists of four main components: *Initialization*, *Tracking and recognition*, *Pose estimation*, *AR scene*.

The *Initialization* component sets required parameters of the camera, the pattern and the 3D object. The *Tracking and recognition* component recognizes the predefined pattern from the image captured by the camera, identifies the geometric features of the pattern and the screen coordinates for future calculation. This step uses SURF method from the *Emgu.CV* software library. The *Pose estimation* component calculates the transformation matrix needed for the establishment of the three-dimensional coordinates on the pattern. *AR scene* component displays the virtual model in the real world.

Two additional software packages were used for implementation of this system:

- NyARToolkit – contains the core of the augmented reality construction and also the implementation of mathematical calculations that are used for determination of the pattern position (see Fig. 2, section “*Pose estimation*”).
- Emgu.CV – is a software library. It contains the SURF method implementation for the detection of patterns in the image (see Fig. 2, section “*Tracking and recognition*”).

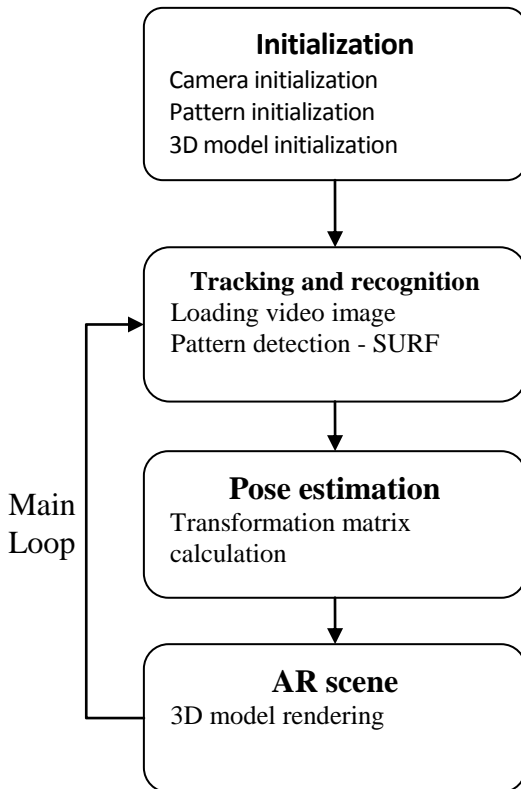


Fig. 2. The architecture of the semi-markerless AR system

A. Tracking and recognition

For recognition of predefined patterns with texture a method called SURF (Speeded Up Robust Features) is used. This method describes the image by using descriptors. Description with the descriptors generated by this method is invariant to rotation and camera distance from the object being described. Points of interest obtained by this method are shown in Fig.4. Detailed description of this method is presented in [5].

B. Pose estimation

Next step after pattern recognition is determination of the pattern's exact position in 3D space. For the calculation itself it is necessary to know the projection matrix, which is obtained by camera calibration [6]. The most important part of calculation is to obtain a transformation matrix that determines the location of the 3D virtual graphic object into 3D space.

The determination of transformation matrix parameters is needed for correct placing of virtual model into the real world. The relationship (Fig.3) between the pattern coordinates and the camera coordinates is described as (1), (2) [6]:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} & V_{13} & W_x \\ V_{21} & V_{22} & V_{23} & W_y \\ V_{31} & V_{32} & V_{33} & W_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} V_{3 \times 3} & W_{3 \times 1} \\ 000 & 1 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} = T_{cm} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} \quad (2)$$

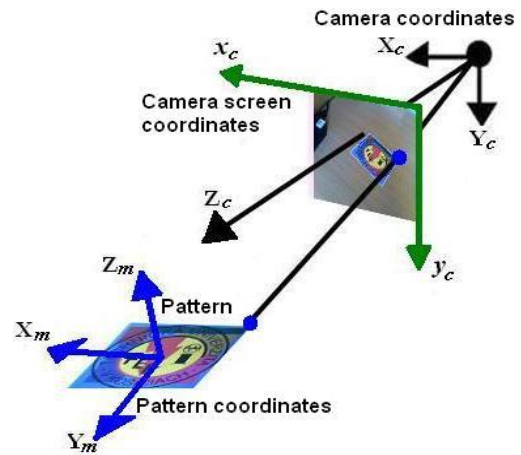


Fig. 3. The relationship between pattern coordinates and the camera coordinates

The matrix T_{cm} (transformation from pattern coordinates to camera coordinates) is obtained by analyzing the input image. This transformation matrix consists from the rotation matrix ($V_{3 \times 3}$) and translation matrix ($W_{3 \times 1}$).

Two parallel patterns edges (margins) are reflected in the image. Coordinates of these edges correspond to the equations of lines (3) [6]:

$$\begin{aligned} l_1 : a_1x + b_1y + c_1 &= 0 \\ l_2 : a_2x + b_2y + c_2 &= 0 \end{aligned} \quad (3)$$

Determination of the line parameters can be calculated in several ways. One of them is calculation of parameters based on the knowledge of at least 2 points that lie on this line. Because the pattern has square or rectangle shape, we can obtain coordinates of its four vertices in the screen coordinate system. These coordinates are obtained using the SURF method after pattern recognition in the video image. Denote the pattern as a rectangle $ABCD$ (Fig.4). Edges AB and CD are parallel. Corresponding equations for these edges are equations of lines l_1 and l_2 (3). Also, the edges BC and DA are parallel and their equations are l_3 and l_4 .

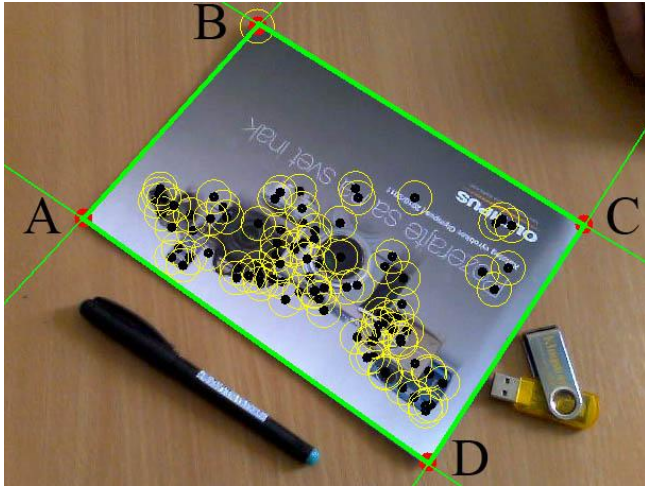


Fig. 4. Rectangle $ABCD$ and points of interest obtained by SURF method

Determination of line parameters for l_1 follows:

1. Finding of direction vector line (3 – 5):

$$\vec{u} = |AB|, A[a_1, a_2], B[b_1, b_2] \quad (3)$$

$$u_1 = b_1 - a_1; u_2 = b_2 - a_2 \quad (4)$$

$$\vec{u} = (u_1, u_2) \quad (5)$$

2. Determination of the vector that is perpendicular to it (6):

$$\vec{n} = (u_2, -u_1) \quad (6)$$

3. Substitution of the values into the general equation of line $ax + by + c = 0$ (7):

$$u_2x - u_1y + c = 0 \quad (7)$$

4. Substitution of the coordinate values x and y for a known point that lies on a line such as coordinates of point B and computation of the parameter c .

In a similar way the general equations of lines l_2 , l_3 , l_4 are obtained.

Next step is to calculate the rotation and translation part of the transformation matrix. This procedure is identical to the procedure presented in [6].

C. AR scene

The system is implemented on MS Windows platform. A 3D model in DirectX format is used as example object for semi-markerless augmented reality implementation, because DirectX libraries are used for rendering (see Fig.2 section “AR scene”).

The result of a semi-markerless augmented reality system as output of the created application is shown in Fig. 5.



Fig. 5. Semi-markerless AR system. The virtual model is displayed in the real world.

III. CONCLUSION

In this paper, a semi-markerless augmented reality application was presented, which was implemented at DCI FEEI TU of Košice. Advantage of this application is that the user does not need to print a classic mark (black square on a white background). The user could use any pattern of square or rectangular shape (e.g. photo, front cover book, the picture in a magazine, etc.), which he has in his surroundings. The application uses the SURF method to detect patterns. This method is relatively difficult to calculate, therefore, one should expect delays in the augmented reality display. This is a disadvantage compared to the classical augmented reality (marker augmented reality) applications.

Future work will be focused on improvement of calculation times needed for the pattern detection.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] R. Azuma, “Tracking Requirements for Augmented Reality”, Communications of the ACM Vol. 36, No. 7, 1993, pp. 50-51.
- [2] P. Milgram; F. Kishino, “A Taxonomy of Mixed Reality Visual Displays”, IEICE Transactions on Information Systems, Vol. E77-D, No. 12, 1994, pp. 1321-1329.
- [3] F. Hrozek, B. Sobota, Cs. Szabó, Š. Korečko, M. Varga, P. Ivančák, “Augmented reality application in parallel computing system”, 7th International Workshop on Grid Computing for Complex Problems,

- Bratislava, Slovakia, 24 - 26 October 2011, Bratislava, Slovakia, Ústav Informatiky SAV, 2011, pp. 118-125, 978-80-970145-5-1
- [4] B. Sobota, R. Janošo, "3D interface Based On Augmented Reality In Client Server Environment", Journal of information, control and management systems, Vol. 8, No. 3, 2010, pp. 247-256, ISSN 1336-1716 .
 - [5] H. Bay, T. Tuytelaars, Luc Van Gool, "Speed Up Robust Features", European Conference on Computer Vision, 1:404-417, 2006.
 - [6] H. Kato, M. Billinghurst, "Marker Tracking and HMD Calibration for a Video based Augmented Reality Conferencing System", In *iwat*, page 85, IEEE Computer Society, 1999.
 - [7] Cs. Szabó, Š. Korečko, B. Sobota, "Data Processing for Virtual Reality," In: Advances in Robotics and Virtual Reality, Intelligent Systems Reference Library Vol. 26, Springer Berlin-Heidelberg, 2012, pp. 333-361, ISBN 978-3-642-23362-3, ISSN 1868-4394.

Material flow in Flexible Assembly Company

¹Ján ILKOVIČ (3rd year), ²Peter PAPCUN (1st year)

Supervisor: ³Ján JADLOVSKÝ

^{1,2,3}Department of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹jan.ilkovic@tuke.sk, ²peter.papcun@tuke.sk, ³jan.jadlovsky@tuke.sk

Abstract— This article describe material flow view to Flexible Assembly Company. The main task is to describe material flow at this system and create the procedure to create a system simulation model. Article contains material flow verbal description and then material flow diagram and at the end the draft of simulation model created in CPNTools environment.

Keywords— flexible assembly system, model, simulation, coloured Petri Net, CPNTools.

I. INTRODUCTION

Over the last years flexible manufacturing systems became a very usual way in solution of all kinds of industry and manufacturing. Growth of production caused interest of increase production efficiency by reason of price reduction per produced goods and also its production time reduction.

It is uneconomic to establish a complex model system or create the required system only for the purpose of testing the system state. The system efficiency increasing and predicting its behaviour can be performed on an existing system, but much cheaper and to create a system model and perform debugging at this model.

This aspect opens the doors for the research in the area of manufacturing system modelling and simulation.

Modelling is a process that studies and analyzes quantitative aspects of a real-world problem. The modelling process results in a model, where mathematics is used to describe the nature of the problem and in many cases a proposed solution.

We had focused our research to a discrete processes and systems, to its modelling and analyzing. The head reason of this our decision was affected by the aspect of system that we wish to research. We look at the system in terms of material flow.

The fundamental principle on which the discrete process model is based is the assumption that all mechanical processes occur not only in discrete space but also in discrete time. Thus, a process is seen as a transition of the system from one state to another at every step in the discontinuous time. These transitions from one state to another during a step in time – the process – are controlled by transition conditionals [5].

Mathematical model is a base requirement for creating a computer interpretable model of knowledge or standard specifications about a kind of process and about relevant facilities of the system.

When we were looking for the best mathematical tool for an implementation our modelling procedure to the system, we

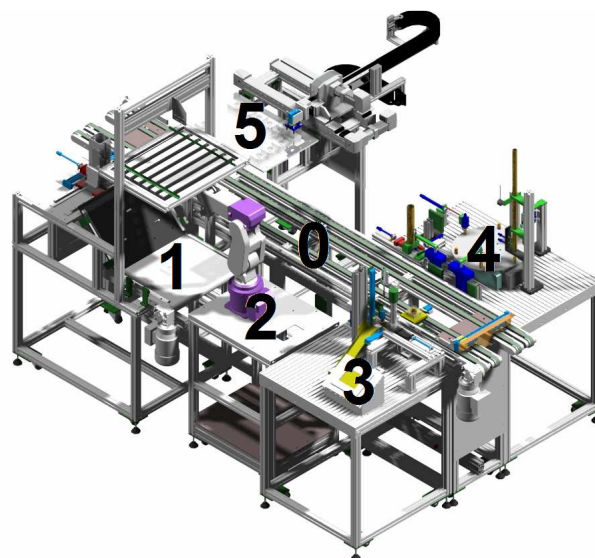


Fig. 1. Graphic design of the assembly line with the numbered posts.

have to analyze our requirements for the tool first. Ability to modelling the parallel systems was main attribute for the searching. Another important option we were looking for was possibility to make discrete event mathematical model with the possibility of computer implementation. As follows we have found Petri Nets with its modification: Petri Nets with modification in Coloured Petri Net. Coloured Petri Net (CPN) is an mathematical and graphical tool supported by wide simulation software base (Matlab, CPNTools,...) we could use for our implementation.

II. PETRI NETS

Petri nets, or place-transition nets, are classical models of concurrency, non-determinism, and control flow, first proposed by Carl Adam Petri in 1962. Petri nets are bipartite graphs and provide an elegant and mathematically rigorous modelling framework for discrete event dynamically systems.

Definition:

A Petri net is a four-tuple (P, T, IN, OUT) where:

- $P = \{p_1, p_2, p_3 \dots p_m\}$ is an set of places;
- $T = \{t_1, t_2, t_3 \dots t_n\}$ is an set of transitions;
- $IN : (P \times T)$ is an input function that defines directed arcs from places to transitions;
- $OUT : (T \times P)$ is an output function that defines directed arcs form transitions to places;

Graphically places are represented by circles and transitions are represented by horizontal or vertical bars interconnected by arcs represented by arrows. Places of Petri nets usually represent states or resources in the system while transitions model the activities of the system. [2]

III. COLOURED PETRI NETS

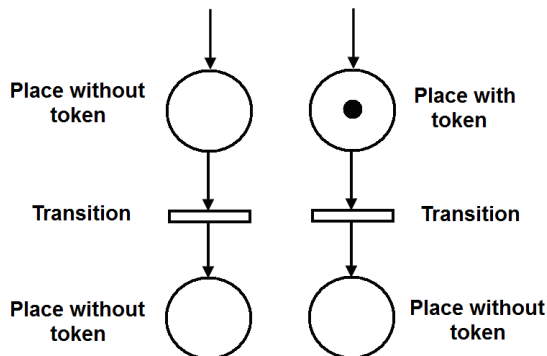


Fig. 2. Left: unmarked PN. Right: marked PN.

Coloured Petri Nets (CPN) is a language for the modelling and validation of systems in which concurrency, communication, and synchronization play a major role. Coloured Petri Nets is a discrete-event modelling language combining Petri nets with the functional programming language Standard ML. Petri nets provide the foundation of the graphical notation and the basic primitives for modelling concurrency, communication, and synchronisation. Standard ML provides the primitives for the definition of data types, describing data manipulation, and for creating compact and parameterisable models. A CPN model of a system is an executable model representing the states of the system and the events (transitions) that can cause the system to change state. The CPN language makes it possible to organise a model as a set of modules, and it includes a time concept for representing the time taken to execute events in the modelled system.

CPN Tools is an industrial-strength computer tool for constructing and analysing CPN models. Using CPN Tools, it is possible to investigate the behaviour of the modelled system using simulation, to verify properties by means of state space methods and model checking, and to conduct simulation-based performance analysis. User interaction with CPN Tools is based on direct manipulation of the graphical representation of the CPN model using interaction techniques, such as tool palettes and marking menus. A license for CPN Tools can be obtained free of charge, also for commercial use. [1]

IV. FLEXIBLE ASSEMBLY COMPANY

Assembly line named Flexible Assembly Company (FAC) was designed and completed at the department of Cybernetics and artificial intelligence by the led of Doc. Ing. Ján Jadlovský CSc.. Model FAC is located in laboratory V147 at Vysokoškolská 4, Košice. The model and its situation at the department laboratory are shown in Fig.1.

The FAC model serves as support for research and diagnostic of industrial communication networks, the production lines modelling and analysis, but the important 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. In practice

would be better to choose alternative technologies to optimize the reliability and simplify the assembly process, but the model purpose 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.

Students have the opportunity to become acquainted with multiple technologies and solutions, which are used in practice, and all in the one place because of using this configuration. The FAP model obviously does not contain all the technologies used in practice, but the designers decided to use just those which he frequently use in the practice, and which also would graduates work with in their practice in the future. More information about the technologies and design of this model are described in [3].

V. FAC MATERIAL FLOW

FAC containing 4 input stores and 1 output one.

As already mentioned above we look at the system in terms of material flow so the main goal of our research is tracing the flow of material in the assembling process. We bind a unique identifier to the products and intermediate products in the simulation model by using which we can track the movement and state of the goods in any time of the production process in the simulation model. It's necessary to know the system and examined its material flow even before the creation of its model. The following section lists a brief description of the assembly FAC process.

The aim of the FAC assembly process is to make a product that is shown in Fig.3 (B). The resultant product consists of four intermediate parts (Fig.3 (A)). These are assembled together by the assembly process in such way that the first part of product is placed on pallette and proceed through the process posts and the intermediate products are progressively mounted together until a resultant product is completed. The product is saved to the output store after assembling finished and there is waiting for order finalization. The entire assembly process can be divided into five positions – posts (Fig.1). 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, where the entire assembly operation begins. The synchronization of individual activities

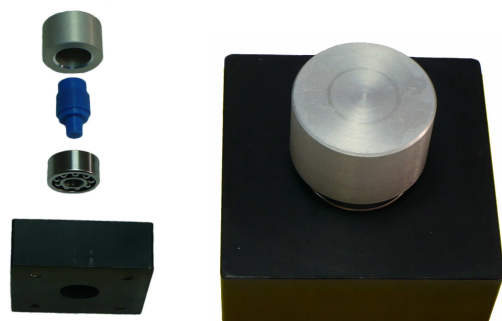


Fig. 3. Individual parts of the product in the order they are assembled (A). The resulting product (B).

so that there are no collisions in the production process is matter of course. 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 whole order. Products are

transferred between the various posts by using a special palette. This palette is not part of the finished product and hence will not be placed in output store. The pallets are situated on a central belt conveyor all time and they are waiting in the queue at the beginning of the assembly process or move product between the posts assembly line.

A. Post 0 – Central belt conveyors

There are two belt conveyors situated in the middle of the FAC. These conveyors serve the transfer of products at each assembly line posts.

B. Post 1 – Ejecting the blocks and camera system

On the Post 1 (Fig.4) starts the assembly process. This post contains a input 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 the serviser to sloping surface. Subsequently, randomly rotated block gets on a wide white belt conveyor located behind the sloping surface. There is also an industrial camera above the belt conveyor.

C. Post 2 – Robotic arm

The block is displaced from the first post (from the belt conveyor) to the central belt conveyor by the robotics arm Mitsubishi RV-2SDB. The robot uses the information from the camera system about the block position.

D. Post 3 – Inserting bearings

After the previous activities is the block transferred by the palette and central belt conveyor to the next post. This post performs inserting a bearing to the product. Bearing, however, must be the correct dimensions. So bearings are picked from a serviser, measured and then inserted into the block. And these activities are carried out in the post 3 (Fig.4).

E. Post 4 – Inserting plastic shafts and metal hats.

Final assembly work is carried out at the post 4 (Fig.4). A plastic shafts and metal caps have to be inserted into the imbedded bearings in the plastic block.

This post consists of 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. After the inserting components in the correct order to the block with bearing is the assembling process finished and proceed to the final post 5.

F. Post 5 – Warehouse

Product stops at the last post and triaxial manipulator moves product from the pallet to its position in the warehouse. 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.

Exporting the finished products from warehouse performs operator by hand.

VI. PROCEDURE FOR MODELLING

A. Preparation of model creating

At the beginning of the modelling process it is necessary to know examined system in terms of material flow. First we have to define system inputs and outputs with its properties. Input and output properties are useful for consecutive declaration of inputs and outputs in the simulation model. The properties will be implemented in the model as parameters of item that will represent input or output.

Next step is to define others system elements that will be used for the material flow model. It's necessary to create a list of system events and states. After that we can connect on each other according to rule: state-event-state. Detailed search and correct describe relations among them ensure success in model creation and in its informative capability.

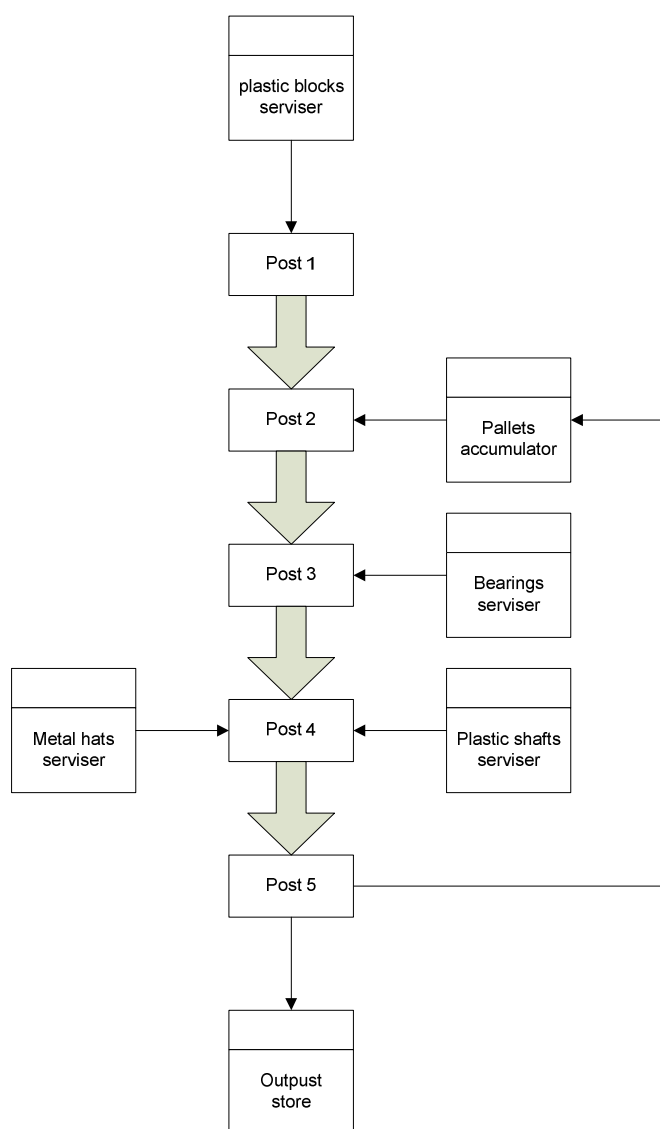


Fig. 4. FAC material flow diagram.

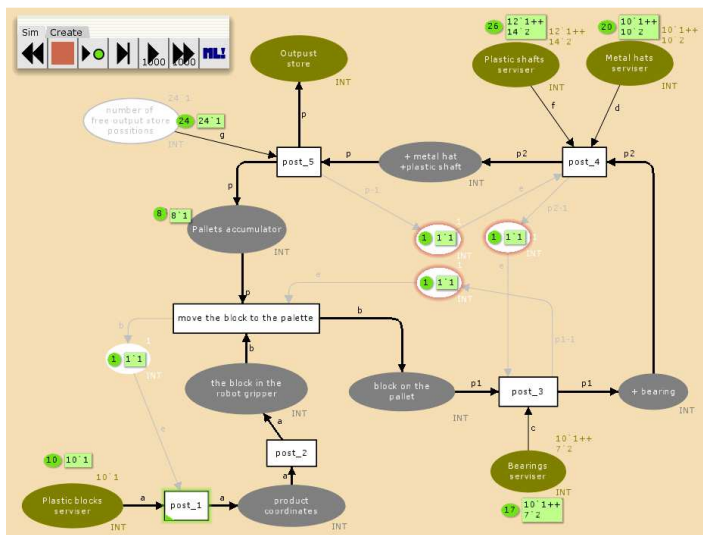


Fig. 5. CPNTools system model – configuration before the simulation starts.

B. Procedure for creating a FAC model

In this section we are going to describe a procedure how to create a simulation model of surveyed system using the Petri Nets. This model is created in the simulation environment CPNTools. We chose this environment for its wide range possibilities for coloured Petri nets modelling and possibility to implement the time variable in the model.

Another reason for our choice was the fact that this application is free downloadable and does not require purchasing a license and therefore whoever could easily download it and use. CPNTools beginner user has to acquaint him with the work at this new environment. There is a possibility to study the manual for work with this environment directly on its official website [4]. And so after a short time, the user can create, save and run modelled system by him and also debug it by running the simulation and its stepping.

We created a simulation CPN FAC model where we can see input and output stores (green ellipses), events (rectangles), system states (grey ellipses) and additive states. We attached figures; first one with system initial conditions [Fig.5] and the state of system at the end of simulation [Fig.6].

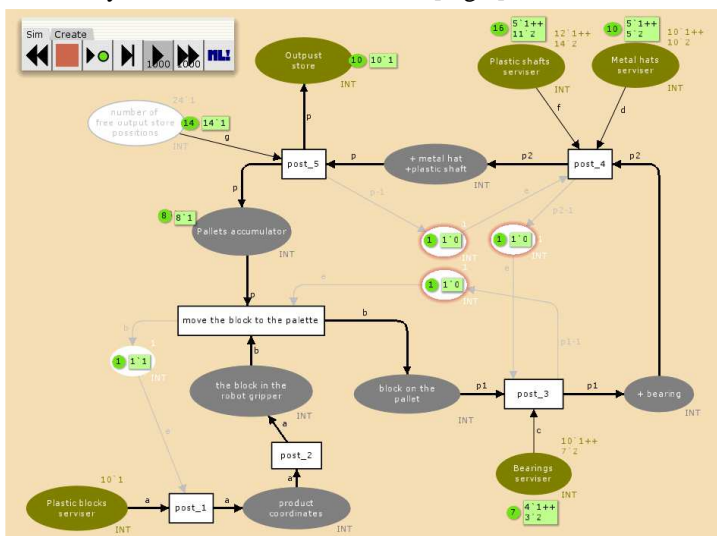


Fig. 6. CPNTools system model – state at the end of simulation.

VII. CONCLUSION

There is ability to use the mentioned procedure in general design for simulate a material flow for much discrete processes of assembly or manufacturing systems. The advantage of this procedure is that there's no requirement to build a system physically, but we can simulate the behaviour of the system, its initial and final state of material.

The CPNTools provides to run the simulation and there is also very useful option: to step the simulation, so we are able to examine the behavior of the system step by step. There is also possibility to add a time element to the model that provides another option for material flow analysis.

ACKNOWLEDGMENT

This contribution is the results of the Scientific Grant Agency of the Slovak Republic and of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030), supported by the Research and Development Operational Program funded by the ERDF.

REFERENCES

- [1] J. Kurt, K. Lars, L. Wells, "Coloured Petri Nets and CPN Tools for modelling and validation of concurrent systems", in International Journal on Software Tools for Technology Transfer, vol. 3, pringer Berlin / Heidelberg, 2007, pp. 213-254.
- [2] J. Ilkovič, "Modelovanie a diagnostika manipulačných a robotizovaných systémov", Písomná časť k dizertačnej skúške, Košice, 2010.
- [3] J. Ilkovič, M. Čopík, T. Karol, "Proposal control of manipulators controlled via technological network DeviceNet", In SCYR: 11th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice: Technická univerzita, 2011, pp. 335-338, ISBN 978-80-553-0644-5.
- [4] CPN tools - <http://cpntools.org/>
- [5] K. Rainio, "Discrete process model", University of Helsinki, ISBN 978-952-10-4547-9.

Mobile Robot Navigation using Adaptive Fuzzy Cognitive Maps

¹Martin Paľa (1st year), ²Ladislav Miženko (2nd year *MSc.)
Supervisor: ³Peter Sinčák

^{2,3}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹martin.pala@tuke.sk, ²ladislav.mizenko@student.tuke.sk

Abstract— This paper shows one of the many possible applications of fuzzy cognitive maps in mobile robot navigation. The description and explanation of the method of image recognition, calculation and execution of efficient intervening cause obtained using fuzzy cognitive map and parametric adaptation of fuzzy cognitive maps are provided in this work. The main purpose is to scheme out a vehicle capable to avoid the obstacles using single camera sensor. The first part of this paper is dedicated to the architecture and hardware configuration of the mobile robot. In the following section the image recognition and fuzzy cognitive map control system is described. The third part of this paper describes the proposed adaptation mechanism of FCM. The last sections are dedicated to the conducted experiments and conclusion of the project and the possible improvements considered for the future work.

Keywords—Fuzzy cognitive map, genetic algorithm, image recognition

I. INTRODUCTION

Navigation is the art of getting from one place to another, safely and efficiently. The means of navigation have changed dramatically over the past decades. From celestial navigation to GPS the methods used for navigation have been upgraded, the share of human factor in navigation has been decreased and it is pushing the human more and more to the role of the observer. This goal can be reached by applying suitable methods of artificial intelligence [1]. The robot navigation problem can be split into two main areas of research: Local navigation and global navigation. The local navigation problem deals with navigation on the scale of a few meters, where the main problem is obstacle. And the global navigation problem deals with navigation on a larger scale in which the robot cannot observe the goal state from its initial position [2].

The main topic of this work is to design and build a vehicle capable of autonomous local navigation and obstacle avoidance in an unknown environment using only one camera as a sensor and fuzzy cognitive map (FCM) as a reasoning component for the movement control.

II. PROBLEM DEFINITION AND ANALYSIS

The task of the autonomous robot navigation using adaptive FCM can be split into two partial tasks. The first one deal

with the problem of acquiring the information about the environment utilizing the sensors of the robot, localizing a robot, obstacles and the target destination and calculating the output data for the robot actuators using non-adaptive FCM generated by an expert. The second partial task is focused on getting and collecting the training data and the following adaptation of the parameters of FCM in order to increase the reliability and accuracy of the whole autonomous mobile robot navigation process. The solution and realization of these two partial tasks depends on the properties and the design data of the mobile robot.

A. Architecture of the autonomous vehicle control and hardware configuration

The architecture of the vehicle control consists of two main parts: Lego NXT vehicle and computer. The vehicle is built by the Lego NXT Mindstorms parts and IP camera as the only sensor used in this mobile robot. The robot built for the experiments is two-wheel, two-motor differential drive vehicle with zero-turn radius. The vehicle is equipped with two NXT motors on both sides for independent side wheel speed control. The structure of the experimental vehicle is similar to Pioneer 2-DX robot. The computer is not a part of the vehicle but is used to gather the data from the IP camera using WIFI communication interface. Then the data are evaluated and sent as instructions back to Lego NXT brick via Bluetooth communication interface. The following picture shows how the communication and data evaluation works.

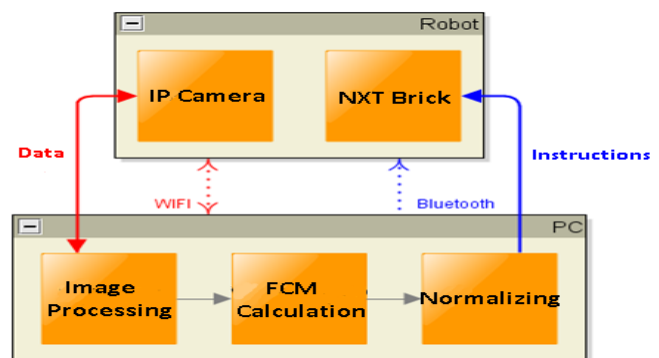


Fig. 1 The control architecture of the proposed mobile robot

The two main elements of the mobile robot, IP camera and NXT brick communicates independently with the workstation.

The blue lines represent Bluetooth communication between NXT brick and computer. On the other side the red lines represent WIFI communication between workstation and IP camera. The process of one mobile robot control loop can be characterized by the following steps. A request code for a picture frame is sent to the IP camera in the first step. In case of successful process of the request, the actual picture frame is sent to workstation via WIFI communication interface. The incoming data are preprocessed, identified and forwarded to the image processing (IP) agent. In this step the environment close to the robot is characterized. Image recognition agent identifies a relative obstacles and target position to the robot position. Finally, this information is used as the input for the calculation of the state of fuzzy cognitive map, which is then transformed to the instructions for the mobile robots actuators. The process of image processing and evaluation of the FCM is described in detail in the following chapter.

III. IMAGE PROCESSING AND FCM EVALUATION

Image processing module of the mobile robot software equipment is not only responsible for target position identification and obstacle detection. It also deals with the problem of the definition of the passable area for the mobile robot. These three features of the environment have been defined for the experimental purposes as follows: Target is marked with red color. The passable area is marked with green or white color and the obstacles are marked with the color different from white, green or red. So every object placed in the experimental arena which is not marked as a target or passable area (floor area) represents the obstacle.

The working cycle of the proposed image processing module consists of three main subroutines executed in sequence: Image preprocessing, image segmentation and image evaluation and output normalization. During the first stage, the image resolution and color depth are lowered, while the main features of the image are preserved. Amended image is then segmented using simple color thresholding method and partial segments are divided into three classes. Each class represents obstacle, passable area or target. The example of the image segmentation used for this project is illustrated in the following picture [3,4].



Fig. 2 Example of segmented image generated in the second phase of IP

Although image in this format is human friendly, it is still difficult for a computer to determine the exact positions of the objects. The image is transformed from bitmap representation to matrix representation so the computational methods are available for further information processing in next stage. By the image to matrix transformation all the redundant information are eliminated and the exact information about the objects can be determined using computational methods.

The following picture shows the example of segmented image in matrix representation form. The image in the proposed system of the object identification is divided into two horizontal parts and three vertical parts, six parts in total.

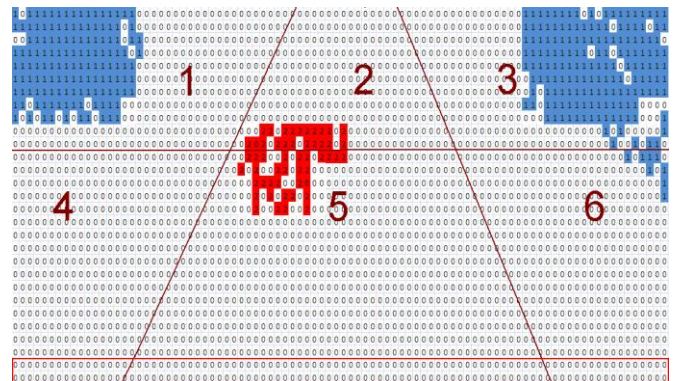


Fig. 3 Example of segmented image using matrix representation

Objects located in the upper part of picture are classified as objects further from the mobile robot. On the other hand objects located in the lower part of the picture are closer to the robot. Objects in sector one and sector four are to the left of the robot. Objects located in sectors two and five are in front of the robot and objects in sectors three and six are to the right of the robot. The special part of the image is below the red line. It is some kind of critical area and all the objects located in this area are critically close to the robot. All these six sectors are evaluated using gradient weight method and each sector is represented by one number which contains the information about measure of the object presence in that sector. The output is then normalized and grouped into one vector which is used as an input for the fuzzy cognitive map.

$$P = [0.21, 0.027, 0.0, 0.02, 0.1, 0.0, 1, 0] \quad (1)$$

The first six elements of the vector represents the measure of the obstacle presence in the following format: obstacle to the left far from the robot, obstacle far in the front, obstacle to the right far from the robot, obstacle to the left close to robot, obstacle close in the front, obstacle to the right close to the robot. The next six elements of the vector represent target position in the same entity sequence. The vector in this format represents a relation between symbolical and numerical values. In the next phase the vector described above is used as an input for the fuzzy cognitive map designed by an expert. The proposed FCM is illustrated in the picture 4.

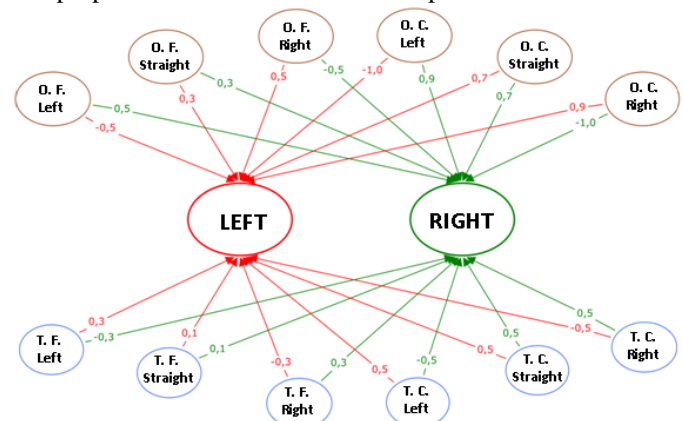


Fig. 4 FCM designed by an expert. (O stands for obstacle, T stands for target, F stands for far and C means close).

FCM is transformed to matrix representation as well, so the computational methods can be used to calculate the output. The output values are defuzzified and normalized [5]. In the last phase of the mobile robot control working cycle, these values are sent to NXT brick as the instructions for the actuators. The mobile robot is capable of autonomous navigation using any FCM with a predefined structure. Now the FCM parameters can be amended to tune up the precision of the control. This can be done again by an expert by conducting more experiments and manually change the parameters or it can be done automatically using of the smart methods. In this paper we present the automatic FCM parameters tuner based on genetic algorithm using data from manual teleoperation. In order to achieve this goal the adaptation module needs to be added to the primary mobile robot control architecture.

IV. FCM ADAPTATION USING GENETIC ALGORITHM

As the FCM structure is dependent on the previous modules used for image processing, the main structure of FCM cannot be changed. The proposed adaptation method uses a genetic algorithm to amend the parameters, causal weights of the FCM. As the adaptation itself needs a strong computation power, this process cannot be run in a real-time. For this purpose the training data collector has been designed to store the control data captured from the human control interface. These data are later used in adaptation process. The general FCM adaptation architecture is illustrated in the following picture.

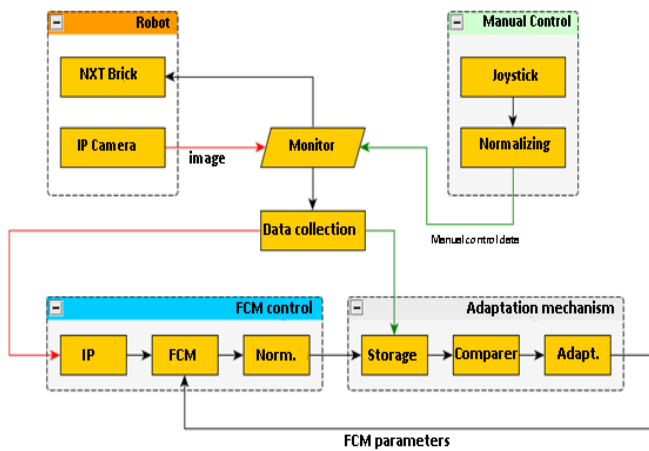


Fig. 5 General architecture of FCM adaptation mechanism

The adaptation of the FCM consists of two main tasks. The training data collection and adaptation process utilizing GA. The training data collection process runs in a real time during the manual control of the mobile robot by an operator. Monitoring system captures the control data and image from the IP camera every predefined unit of time. The adaptation process is driven by an error. Error is calculated as the difference between the FCM output and manual control data. The FCM output is calculated using the stored image data in the training data file and paired with the manual control data from the human operator. The role of the genetic algorithm is to change the parameters of the FCM so the FCM output data values are close to manual control data over the whole set of training data, so the error between autonomous FCM control output and manual operator control is minimal. The proposed genetic algorithm used for the FCM adaptation consists of the

standard GA blocks. Each population individual represents all the parameters of FCM. Each gene of the chromosome is represented as a real number in interval $<-1,1>$ and represents each FCM parameter. The process of the fitness value calculation consists of two phases. In the first phase the chromosome is transferred to FCM and in the second phase the mean error is calculated over the whole training data set. The error is calculated as a difference between the outputs generated for right and left motors of the mobile robot calculated using FCM using the stored data of images in the training data file and the control data generated by a human operator during the manual teleoperation process, also stored in the training data file. The mean error then represents the fitness of the individual. The fitness value is calculated using the following formula:

$$E_a = \frac{\sum_{i=0}^n (|L_{F_i} - L_{H_i}| + |R_{F_i} - R_{H_i}|)}{n} \quad (2)$$

Variable L_F represents the output value calculated using FCM for left motor. Variable L_H represents the output value captured from human operator for the left motor of the mobile robot. Variable R_F is the output value calculated by the FCM for the right motor and R_H is an output value generated by a human operator during the manual control process of the mobile robot. The number of the training data patterns is represented by the variable n . After the evaluation of all of the individuals in the population of GA, the algorithm continues in a standard way. The adaptation process ends when the error value is zero or the GA reaches the maximum defined generations.

V. EXPERIMENTS

The autonomous control system of the mobile robot and the adaptation extension has been tested in various conditions and different situations of obstacles and target placement in the testing area. During the experiments, the standard FCM designed by an expert and adapted FCM by the proposed genetic algorithm has been compared. The experiments showed that the movement of the mobile robot using FCM designed by the expert is not smooth and robot is changing the direction very often. In addition the accuracy of the system is not very high. In some cases the robot lost visual contact with the target and got lost. On the other hand after conducting experiments with learning system, the parameters of the FCM has been changed and positively influenced the performance of the mobile robot navigation system. The best results have been achieved using the following parameters of the adaptation system:

- Training data collection interval 200ms
- Training data patterns 100
- GA population size 40
- Number of generations 20
- Selection type random
- Mutation rate 2%

It was not possible to increase the number of generations or population size as the adaptation process with these parameters took about one hour. All the experiments are available for download.

VI. CONCLUSION

The Proposed obstacle avoidance and navigation system has been tested in the environment with the various numbers of obstacles and obstacle alignments with a static or moving target destination. The same experiments were conducted for the mobile robot using FCM designed by an expert and FCM with adapted parameters by the proposed adaptation system. The autonomous vehicle using expert only designed FCM smoothly detected and avoided single obstacle. In the environment with the multiple obstacles the mobile robot with the same configuration also proved that it is capable of multiple obstacle avoidance. In some cases the robot had to use the critical brake system to not to drive in the obstacle, but after a few new tries it successfully avoided all the obstacles and reached the given destination. However the behavior of the mobile robot was not always predictable and well-balanced. In some cases the movement of the robot was not so smooth and the vehicle was changing the direction very often even if it was not necessary needed. The proposed FCM adaptation system solved all these problems. Even with a small amount of training data collected from the operator during the manual robot control, the proposed adaptation system was able to positively amend the parameters of the FCM and tune up the control process. The experiments proved that the proposed method of FCM casual weight adaptation is able to improve the control process of the mobile robot and that the systems capable of learning from teleoperation might have a strong impact on the future robotic systems.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] Martin Paľa, Ján Vaščák: Mobile Robot Navigation Using Fuzzy Cognitive: Electrical Engineering and Informatics 2: Proceeding of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice. - Košice: TU, FEI, 2011 S. 309-312. - ISBN 978-80-553-0611-7
- [2] M. A. Batalin, G. S. Sukhatme: Mobile Robot Navigation using a Sensor Network, In IEEE International Conference on Robotics and Automation, pp. 636-642, New Orleans, LA, April 26 - May 1, 2004.
- [3] SKARBK, Wladyslaw – KOSCHAN, Andreas: Colour Image Segmentation: A Survey. Berlin: Technical University of Berlin, 1994. 81 s.
- [4] SOWMYA, B. - SHEELARANI, B.: Colour Image Segmentation Using Soft Computing Techniques. In: International Journal of Soft Computing Applications ISSN: 1453-2277, Issue 4 (2009), pp.69-80.
- [5] PAPAGEORGIOU, E. – MARKINOS, A. – GEMPTOS, T.: Application of fuzzy cognitive maps for cotton yield management in precision farming, Expert Systems with Applications 36 (2009) 12399.

Multimedia solution for live streaming of various types of educational content

¹*Dávid CYMBALÁK (1st year)*

Supervisor: ²František JAKAB

^{1,2}Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹david.cymbalak@cnl.sk, ²frantisek.jakab@cnl.sk

Abstract— This paper describes the concept of multimedia solution for realizing live streaming of various types of content especially educational content. As a source of streaming can be used video camera, video file, audio, images, presentations, desktop capture, etc. Those sources are mixed in real time in video mixer software to one resulting output of video stream. Output of video stream is transferred to virtual camera device, which is set as input in encoder. Encoding of video must be realized in proper bit-rate depending on type of streaming source. Encoded video is streamed via media server to flash player located in web interface. Mentioned functionality enriches the teaching process and brings ability to deliver combination of almost every educational material to remote student in form of video streaming. For remote students it brings opportunity to have more understandable and more attractive educational materials. Result of this work was experimental realization of proposed solution, which was tested by streaming chosen educational sources.

Keywords—Streaming, E-learning, Video mixer, Virtual camera

I. INTRODUCTION

In present the topic of streaming technology is quite extensive and thus more and more widespread in commercial services and also in the educational field. Distance learning realized by streaming technologies could be in the future as one of the primary sources of education. The aim of this work is to contribute a new concept of streaming solution with ability of mixing sources of learning objects in one live broadcast.

Video is a rich and powerful medium, which is being used in e-learning. It can present educational information in more attractive and more understandable way. It allows students to view learning objects or realistic examples, to see sequences in motion, and to listen to narration. Video streaming is an e-learning technology that transmits a continuous video stream over the Internet as digital codes which are then reinterpreted as moving images to a compatible web browser for instant playback [1]. Video streaming can be live or delayed. Video streaming allows those students attending class to review previous lectures and update notes at their convenience. Students who rely strictly on streaming video can view lectures on their own time, at their own pace, and as

frequently as needed. Students can learn efficiently and report satisfaction with a flexible learning experience [2]. Streaming media formats compared with conventional media formats brings many advantages. These benefits include: media content protection against piracy, controlling the flow of content, selection of allowed recipients or easy content management. For this reason, especially due to affordability, these technologies are largely appearing as an appropriate means for providing educational services through the Internet. This all brings a new dimensions and more effective teaching methods than static text and images.

Delivering of streaming media content is provided without the need to download entire files. As soon as computer starts playing the media, it also starts downloading content from the next sections from streaming server. This process is called buffering, which means that downloading content and playing the parts are going at the same time. Streaming is mostly without affecting the transmission quality, in addition to the initial buffer initialization. [3]

II. SOLUTION DESIGN

A. Expected functionality

Based on observation on common lecture, there can be defined expected functionality of proposed solution in those points:

- Allow live streaming of video or audio from lecture.
- Allow streaming video files, audio files or images.
- Allow stream slides from presentations.
- Allow to stream graphical output from teachers laptop.
- Provide possibility to dynamically change the source of streaming during the live broadcast.
- Record live streaming to file to archive.
- Deliver broadcast to students via web interface.

For satisfying expected requirements, there is need to analyze proper technologies such as streaming technologies, video mixers or web technologies and there is also need to create a interconnections between chosen components in form of newly created elements.

B. Conception of chosen components

Conception of solution is based on creation of virtual camera device (Fig 1.). This device consists of output from software video mixer, where the video is finalized from various types of sources. External output from virtual device is encoded to flash by encoder and is transported via streaming server to flash player located on web page. Simultaneously there is possibility to archive output in file.

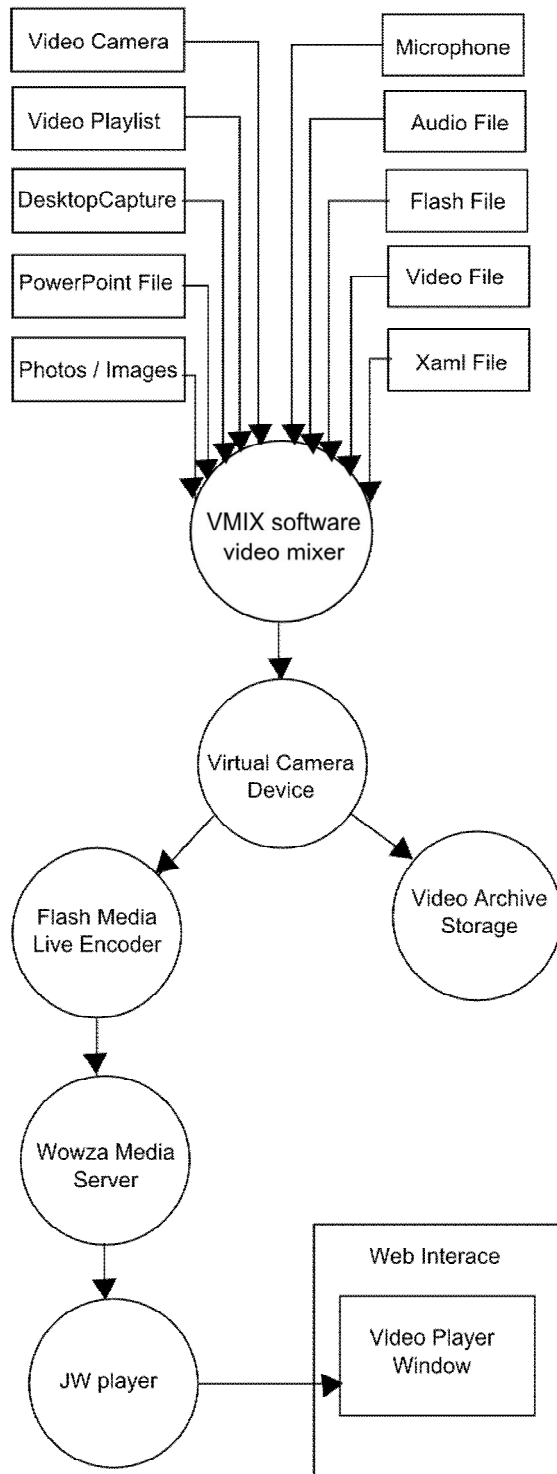


Fig. 1. Design of proposed solution with virtual camera device

Mixing the sources is realized by software video mixer called Vmix, the tool that utilizes the latest advances in computer hardware to provide live HD video mixing, a task previously only possible on expensive dedicated hardware mixers. Encoding the resulting output is utilized by Flash Media Live Encoder, which has various features influencing quality of streaming beginning from selecting the encoding resolution, ending with the choice of filters to determine the bit rate. Transporting encoded video is ensured by Wowza media server, which fully interactive server for streaming multimedia content with full support for H.264 and RTMP protocol. Watching of broadcast is possible via JW player, which is web flash player bringing feature called dynamic bit rate switching, which set the bit rate of video dynamically according the information from user about bandwidth. This feature decides what bit rate has to be provided to best ensure the continuity and quality of play [3].

C. Streaming Bandwidth Based on Type of Video Source

Important thing for ensuring the readability or visibility of streamed materials is quality of video output. Significant impact on the quality of high resolution has the way how the image is recorded and rendered on the screen. Encoding is provided by codec. Codec is compression algorithm, which is used to reduce the file size. Each codec has ability to stream with specific amount bits per second (bit rate), which depends on the level of compression, on the resolution of the video, on the information contained in the transmission or on the number of audio channels. This codec provides encapsulation of video content in video file format container [4,5]. In this solution it is necessary to define appropriate resolution, codec, file container and bit rate for each streaming source. As a compromise between quality and bandwidth it was chosen the HD (1280x720) resolution, flash file format container – FLV and encoding via VP6 codec.

Even though the quality is not the primary feature of streaming technology, it is the way, how can affect quality. When it comes to multimedia transmission through lines with high bandwidth, then can be used appropriate encoding, which ensures higher quality. When it comes to the transmission through lines with lower bandwidth, then it is possible to use the encoding, which will primarily ensures smooth playback. Choosing the appropriate bit rate was subject of testing on various samples. It was created the small amount of live broadcasts with bit rate in range from 500kb/s to 5000kb/s streaming sample video from camera capturing lector, streaming text presentation or streaming desktop capture. Streaming from camera source or video file capturing lector speech and movements had acceptable quality in 3500kb/s, but it depends on count of pixel changes in each frame in video. When camera is capturing person with lack of movements, quality of stream can be acceptable in 2500kb/s in HD resolution. Streaming presentations or desktop capture had better results, because in this case the average count of pixel changes is very small in each frame. The acceptable quality of video starts with 500kb/s or 1000kb/s in case of situation, when presented slides are changing very quickly or capturing desktop is very dynamic.

III. EXPERIMENTAL REALIZATION

Overall solution was tested in form of experimental realization of interconnection between mentioned components. Using PHP, HTML and jQuery UI, there was created the web page containing JW player for watching the encoded output from Wowza mixed by Vmix. It was tested to stream different types of sources in different bit rates. Capturing video from desktop was realized by Vmix desktop capture server from PC located in LAN. The possibility to overlay sources in form of PIP or overlay information banner in XAML file was also realized. (Fig 2,3,4)



Fig. 2. Mixing sources in software video mixer

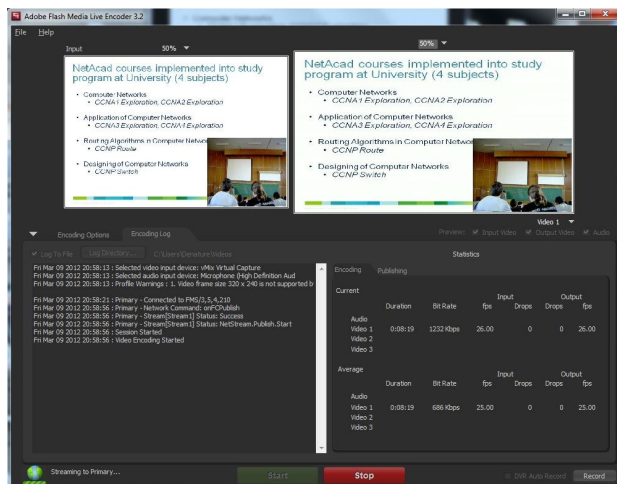


Fig. 3. Encoding virtual camera device in FLME

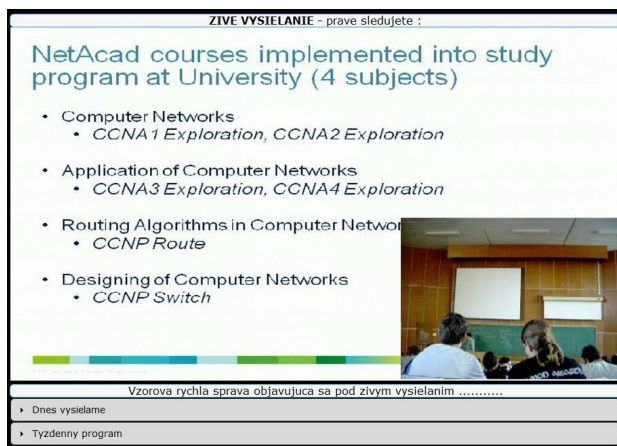


Fig. 4. Watching the encoded output via Wowza server in JW player

IV. CONCLUSION

Proposed solution represents new low-cost alternative for distance learning, where streaming of educational materials is adapted to various type of source content. Delivering videos, presentations or desktop capture in real time to students creates teaching process more effective and more interesting for participants.

Future work on this solution will be concentrated on implementing comprehensive web portal, which will have not only described functionality, but also functionality such as automate publishing and post processing of archived videos or ability to dynamically upload educational materials during live streaming.

It could be mentioned, that educational multimedia technologies have also many disadvantages in form of high costs, technical maintenance or experienced operation. But also there is a loss of personal contact between teachers and students. It could have a negative impact on the development of student's communication and expressive skills. The one of the biggest problem is to bring these modern technologies in to the practice. Sometimes it represents a barrier, when teachers must continually educate to keep pace with the times. On the other hand, brings many benefits for learning process, which is becoming more efficient and more interesting.

ACKNOWLEDGMENT

This work is the result of the project implementation Development of the Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] Shephard K. *Questioning, promoting and evaluating the use of streaming video to support student learning*. Br J Educ Technol. 2003
- [2] Demetriadis S, Pombortsis A. *e-lectures for flexible learning: a study on their learning efficiency*. Educ Technol Soc. 2007
- [3] Cymbalak D., *Next generation IPTV*, Diploma thesis, Kosice: Technical University of Kosice, 2011.
- [4] O'Driscoll G., *Next generation IPTV services and Technologies*, Hoboken: Wiley, 2008.
- [5] Michalko M., *Creation of videostreamings for mobile devices with required quality*, Ph.D.Thesis,Kosice : Technical University of Kosice, 2010

Nonlinear Control Design of Robot Arm

¹Peter ŠUSTER (3rd year)
Supervisor: ¹Anna JADLOVSKÁ

¹Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹peter.suster@tuke.sk, ¹anna.jadlovsk@tuke.sk

Abstract— The aim of this paper is control algorithm design for a robot arm nonlinear model using an exact input - output feedback linearization method. In this paper is given the nonlinear model of the robot arm with two degree of freedom, which is base of the simulation model, and also a detailed description of the exact input - output feedback linearization method. The proposed control algorithm together with simulation model of the robot arm are implemented into control structure with purpose to track reference trajectory.

Keywords— robot arm, exact feedback linearization method, pole placement method, integrator

I. INTRODUCTION

The industrial robots have become a very important part of the manufacturing facilities in the various fields of industry. The robot arms are used for different kinds of welding, assembling, painting and other industrial applications. Therefore, the high level control is required for these applications, which provide the desired precision of the robot arm. Several approaches were used to control algorithm design for robot arms as robust control [2], or genetic algorithm for optimization of trajectory [3] and etc.

In this paper will be presented application of the exact input - output feedback linearization method for control algorithm design for nonlinear simulation model of robot arm with two degree of freedom with purpose of control on steady states defined by a reference trajectory. The proposed control algorithm together with robot arm simulation model are implemented into control structure, which is then verified in the Matlab/Simulink program language.

II. EXACT INPUT-OUTPUT FEEDBACK LINEARIZATION METHOD FOR MIMO SYSTEM

This part described the exact input - output feedback linearization method for nonlinear dynamic MIMO of n - order with m - number of inputs and outputs. The basic condition for using exact linearization method is nonlinear MIMO system described in the affine form

$$\begin{aligned}\dot{x}(t) &= f(x, t) + g_1(x, t)u_1(t) + \dots + g_m(x, t)u_m(t) \\ y_1(t) &= h_1(x, t) \\ &\dots\end{aligned}\quad (1)$$

$$y_m(t) = h_m(x, t)$$

$$i = 1..m \quad - \text{ } i\text{th inputs}$$

$$j = 1..m \quad - \text{ } j\text{th outputs}$$

where $x(t) \in R^n$ is state vector, $u_i(t)$ is control input, $y_j(t)$ is

system output, $f(x, t)$, $g_i(x, t)$ a $h_j(x, t)$ are smooth nonlinear functions. For better overview, further will not write dependence of variables on the time t .

The essence of the exact input - output feedback linearization method is in finding a input transformation in the shape

$$u_i = \alpha_i(x) + \beta_i(x)v_i \quad (2)$$

where v_i is new input, $\alpha_i(x)$, $\beta_i(x)$ are nonlinear functions, and in defining a state transformation z_i in the shape

$$\hat{z}_i = T_i(x) \quad (3)$$

The linear relationship is then created among outputs y_j and the new inputs v_i , and the interactions are removed between original inputs and outputs. That is advantage of this method, which allows to design control algorithm for each subsystems with input v_i and output y_j and independent of each other by synthesis methods for SISO systems [1].

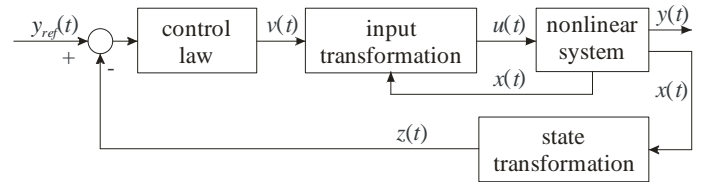


Figure 1 Control structure using exact linearization method

Principle of the exact input - output method feedback linearization method is based on repeatedly derivative of output y_j until input signals appear in the expression of derivation. The Lie derivatives are used for the calculation of individual derivatives of outputs, which are marked as $L_f h$ and $L_{g_i} h$. The first derivative has the form

$$\dot{y}_j = L_f h_j(x) + \sum_{i=1}^m L_{g_i} h_j(x) u_i \quad (4)$$

where:

$$L_f h_j(x) = \frac{\partial h_j}{\partial x} f(x), \quad L_{g_i} h_j(x) = \frac{\partial h_j}{\partial x} g_i(x)$$

If expression $L_{g_i} h_j(x) = 0$ for all i , then the inputs have not appeared in the derivation and is necessary continues derivative of the output y_j . Precondition, the number r_j represents number of derivatives, that is needed, that at least one input has appeared in the derivation $y_j^{r_j}$ i.e.

$L_{g_i} L_f^{r_j-1} h_j(x) \neq 0$ at least for one i , then the resulting derivation has shape

$$y_j^{r_j} = L_f^{r_j} h_j(x) + \sum_{i=1}^m L_{g_i} L_f^{r_j-1} h_j(x) u_i \quad (5)$$

This approach must be done for each output y_j . When the derivation is finished, the result are m equations, which can be written in the form

$$\begin{bmatrix} y_1^{r_1} \\ \dots \\ y_m^{r_m} \end{bmatrix} = \begin{bmatrix} L_f^{r_1} h_1(x) \\ \dots \\ L_f^{r_m} h_m(x) \end{bmatrix} + E(x) \begin{bmatrix} u_1 \\ \dots \\ u_m \end{bmatrix} \quad (6)$$

where $E(x)$ is $m \times m$ matrix of shape

$$E(x) = \begin{bmatrix} L_{g_1} L_f^{r_1-1} h_1 & \dots & L_{g_m} L_f^{r_1-1} h_1 \\ \dots & \dots & \dots \\ L_{g_1} L_f^{r_m-1} h_m & \dots & L_{g_m} L_f^{r_m-1} h_m \end{bmatrix} \quad (7)$$

If matrix $E(x)$ is regular, then it is possible to define the input transformation in the shape

$$\begin{bmatrix} u_1 \\ \dots \\ u_m \end{bmatrix} = -E^{-1}(x) \begin{bmatrix} L_f^{r_1} h_1(x) \\ \dots \\ L_f^{r_m} h_m(x) \end{bmatrix} + E^{-1}(x) \begin{bmatrix} v_1 \\ \dots \\ v_m \end{bmatrix} \quad (8)$$

and state transformation

$$\begin{bmatrix} \hat{z}_1 \\ \dots \\ \hat{z}_m \end{bmatrix} = \begin{bmatrix} T_1(x) \\ \dots \\ T_m(x) \end{bmatrix} = \begin{bmatrix} [h_1(x), L_f h_1(x), \dots, L_f^{r_1-1} h_1(x)]^T \\ \dots \\ [h_m(x), L_f h_m(x), \dots, L_f^{r_m-1} h_m(x)]^T \end{bmatrix} \quad (9)$$

After determination of input transformation (8) and state transformation (9) can be transformed nonlinear system (1) into linear form

$$\dot{z} = Az + bv = \begin{bmatrix} 0 & I_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & 0 & \dots & I_m \\ 0 & \dots & \dots & 0 \end{bmatrix} \begin{bmatrix} \hat{z}_1 \\ \vdots \\ \hat{z}_m \end{bmatrix} + \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \vdots & \vdots \\ 1 & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \quad (10)$$

where I_j are unit matrix of the size $r_j \times r_j$. The linear control algorithm is necessary to propose feedback control law for linear system (10) by linear method of synthesis, to ensure the desired behavior of the nonlinear system (Fig.1).

The relative order r_j exist for each subsystem in the exact input - output feedback linearization method, the resulting relative order is then defined by the sum of them such as

$$r = r_1 + r_2 + \dots + r_m \quad (11)$$

This part described the exact input - output feedback linearization method for the case when relative order r is equals to system order n . [1]

III. SECOND-ORDER DYNAMIC MODEL OF ROBOT ARM

The robot arm model with two degrees of freedom (2DOF) is nonlinear MIMO system with two inputs (voltages for DC motors) and with two outputs (angular position of individual joint) (Fig. 2). The model is divided into two subsystems, *DC motors* subsystem, which consists two DC motors, that convert the input voltage u_j to corresponding motor torque τ_j , and *Robot arm* subsystem, which simulates the movement of robot arm joints (Fig.3).

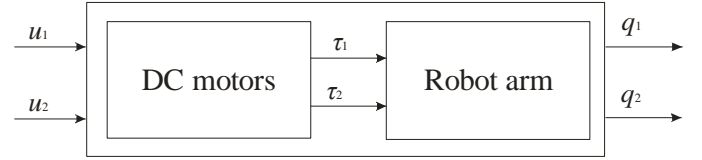


Figure 2 Model of robot arm - subsystems

The mathematical model of *DC motors* subsystem composed of two DC motors, which are described by differential equation of the shape [2]

$$R_j i_j + L_j \frac{di_j}{dt} + K_{b_j} \dot{q}_j = u_j \quad (12)$$

$j = 1, 2$

where physical variables and parameters are:

- i_j - armature current of motor
- u_j - input voltage of motor
- \dot{q}_j - angular position of rotor
- R_j - resistance of armature of motor
- L_j - inductance of armature of motor
- K_{b_j} - EMF constant

The motor output torque is then given

$$\tau_j = K_{t_j} i_j \quad j = 1, 2 \quad (13)$$

where τ_j is motor output torque of j th joint of robot arm and K_{t_j} is torque constant of j th motor. In the further, it is assumed, $K_{b_j} = K_{t_j}$ for both joints of the robot arm.

The general description of the dynamic of the robot arm with k - degree of freedom can be given by motion equation of the shape [4]

$$M(q)\ddot{q} + H(\dot{q}, q) = u \quad (14)$$

where:

- q - angular position vector [$k \times 1$]
- \dot{q} - angular velocity vector [$k \times 1$]
- \ddot{q} - angular acceleration vector [$k \times 1$]
- $M(q)$ - matrix of inertia [$k \times k$]

$H(\dot{q}, q)$ - vector of damping, coriolis, centrifugal and gravitational force [$k \times 1$]

u - input torques vector [$k \times 1$].

In the case 2DOF robot arm (Fig 3), which includes a *Robot arm* subsystem, can be the dynamic of the arm described follows

$$M(q) \begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \end{bmatrix} + \begin{bmatrix} \hat{h}_1 \\ \hat{h}_2 \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} \quad (15)$$

where τ_1, τ_2 are DC motors torque, q_1 and q_2 are angular position of individual joints. The matrix of inertia $M(q)$ and $H(\dot{q}, q)$ vector are in the shape

$$M(q) = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} (m_1 + m_2)l_1^2 & m_2 l_1 l_2 \cos(q_1 - q_2) \\ m_2 l_1 l_2 \cos(q_1 - q_2) & m_2 l_2^2 \end{bmatrix} \quad (16)$$

$$H(\dot{q}, q) = \begin{bmatrix} \hat{h}_1 \\ \hat{h}_2 \end{bmatrix} = \begin{bmatrix} K_{q_1} \dot{q}_1 + m_2 l_1 l_2 \sin(q_1 - q_2) \dot{q}_2 \\ K_{q_2} \dot{q}_2 - m_2 l_1 l_2 \sin(q_1 - q_2) \dot{q}_1 \end{bmatrix} \quad (17)$$

where:

- m_1, m_2 - mass of arms

l_1, l_2 - length of arms
 K_{q1}, K_{q2} - damping constants.

It is assumed that the total mass of the actual arm is concentrated at the end of each of them, as shows on the Fig.3.

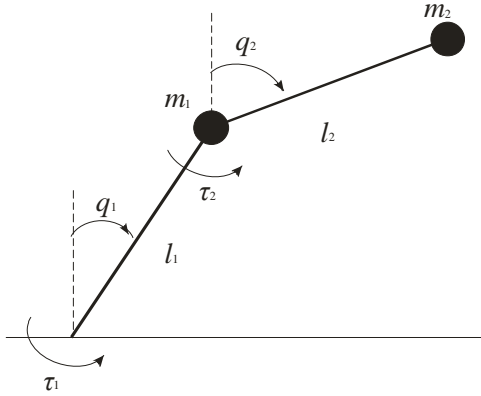


Figure 3 2DOF robot arm

Based on equations (12) to (17) was programmed simulation model of the nonlinear dynamic model 2DOF robot arm using the S - function block in the Matlab/Simulink language. The simulation model includes physical constraints and limits of the real model. The analysis of the simulation model was carried in open loop at defined input voltages u_1 and u_2 (Fig.4).

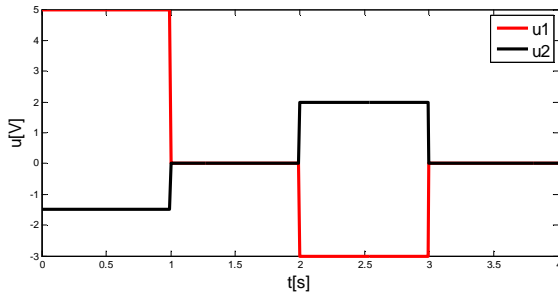


Figure 4 Analysis of 2DOF robot arm simulation model - input

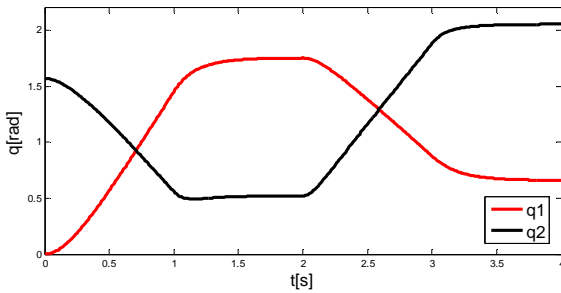


Figure 5 Analysis of 2DOF robot arm simulation model - output

The resulting graph of analysis (Fig.5) shows the possibility of further use of 2DOF robot arm simulation model in the control structure at the testing of the control algorithm using exact input - output feedback linearization method.

IV. DESIGN OF NONLINEAR CONTROL ALGORITHM

The nonlinear dynamic model of the 2DOF robot arm was given in the previous part. This model is described by equations, which contain in itself a smooth nonlinear functions, the main assumption for the use of exact input - output feedback linearization method for the control algorithm design, which will deal in this part. The whole process of the control algorithm design is described in the following steps.

1. step - rewrite of the nonlinear model of the 2DOF robot

arm to affine form (1).

By defining the state vector $x = (x_1, x_2, x_3, x_4, x_5, x_6) = (q_1, \dot{q}_1, i_1, q_2, \dot{q}_2, i_2)$, system input $u = (u_1, u_2)$ and system output $y = (q_1, q_2)$, then the affine form of the 2DOF robot arm nonlinear model has

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \\ \dot{x}_6 \end{bmatrix} = \begin{bmatrix} x_2 \\ -invM_1 H + invM_1 \tau_1 \\ -\frac{R}{L} x_3 - \frac{K_b}{L} x_2 \\ x_5 \\ -invM_2 H + invM_2 \tau_2 \\ -\frac{R}{L} x_6 - \frac{K_b}{L} x_5 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ L \end{bmatrix} u \quad (18)$$

$$y = \begin{bmatrix} x_1 \\ x_4 \end{bmatrix}$$

where

$$invM = \begin{bmatrix} invM_1 \\ invM_2 \end{bmatrix} = \begin{bmatrix} \frac{m_{22}}{m_{11}m_{22} - m_{12}m_{21}} & -\frac{m_{21}}{m_{11}m_{22} - m_{12}m_{21}} \\ -\frac{m_{12}}{m_{11}m_{22} - m_{12}m_{21}} & \frac{m_{11}}{m_{11}m_{22} - m_{12}m_{21}} \end{bmatrix}$$

$$H = H(\dot{q}, q)$$

$$\tau = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} K_t x_3 \\ K_t x_6 \end{bmatrix}.$$

2. step - calculation of the Lie derivatives for individual outputs.

The program module has been programmed for this calculation in Matlab language using symbolic toolbox, that generated the individual derivatives for outputs y_1 and y_2 . After calculation, the result are matrix of the $L_f h(x)$ derivatives and $E(x)$ matrix in the shape

$$L_f h(x) = \begin{bmatrix} L_f^0 h_1(x) & L_f^0 h_2(x) \\ L_f^1 h_1(x) & L_f^1 h_2(x) \\ L_f^2 h_1(x) & L_f^2 h_2(x) \\ L_f^3 h_1(x) & L_f^3 h_2(x) \end{bmatrix} \quad (19)$$

$$E(x) = \begin{bmatrix} L_{g_1} L_f^2 h_1 & L_{g_2} L_f^2 h_1 \\ L_{g_1} L_f^2 h_2 & L_{g_2} L_f^2 h_2 \end{bmatrix}. \quad (20)$$

After determining, that $E(x)$ matrix is regular, it was possible to define the input transformation

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = -E^{-1}(x) \begin{bmatrix} L_f^3 h_1(x) \\ L_f^3 h_2(x) \end{bmatrix} + E^{-1}(x) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad (21)$$

and state transformation

$$\begin{bmatrix} \hat{z}_1 \\ \hat{z}_2 \end{bmatrix} = \begin{bmatrix} T_1(x) \\ T_2(x) \end{bmatrix} = \begin{bmatrix} [h_1(x), L_f^1 h_1(x), L_f^2 h_1(x)]^T \\ [h_2(x), L_f^1 h_2(x), L_f^2 h_2(x)]^T \end{bmatrix}. \quad (22)$$

3. step - transformation of the nonlinear system (18) to linear form (10) and linear control law.

The linear control law was proposed for to ensure the desired behavior of the nonlinear model (18) using pole placement with integrator method in the shape

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = - \begin{bmatrix} K_{1-3} & K_{i1} \\ K_{4-6} & K_{i2} \end{bmatrix} \begin{bmatrix} \int y_{ref1} - y_1 \\ z_{1-3} \\ \int y_{ref2} - y_2 \\ z_{4-6} \end{bmatrix} \quad (23)$$

where:

- z_{1-3}, z_{4-6} - state vectors
- y_{ref1}, y_{ref2} - reference trajectories
- y_1, y_2 - model outputs
- K_{1-3}, K_{4-6} - vectors of gains for individual states
- K_{i1}, K_{i2} - vectors of gains for integrator outputs

Therefore, it was necessary to extend the state description of the linear form (10) into following form

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \\ \int \dot{z}_1 \\ \dot{z}_4 \\ \dot{z}_5 \\ \dot{z}_6 \\ \int \dot{z}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \int z_1 \\ z_4 \\ z_5 \\ z_6 \\ \int z_4 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad (24)$$

4. step - implementation of the proposed control algorithm and his testing in the control structure.

The resulting input transformation (21) and state transformation (22) together with linear control law (23) are implemented into programmed simulation scheme for control of the 2DOF robot arm nonlinear model (Fig.6)

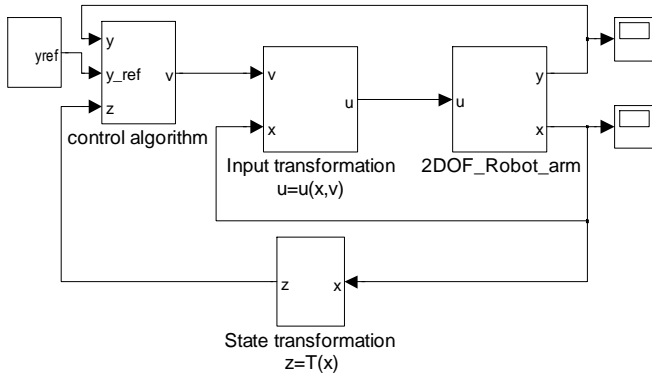


Figure 6 Simulation scheme for control 2DOF robot arm nonlinear model using exact input - output feedback linearization method

The resulting graph of tracking reference trajectory, which is represented the step change between steady state, when using proposed control algorithm with using exact input - output feedback linearization method is on the Fig.7.

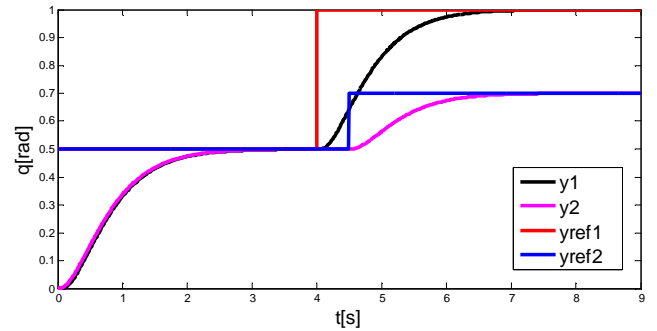


Figure 7 2DOF robot arm nonlinear mode response to track reference trajectory

V. CONCLUSION

This paper presented the control algorithm design for nonlinear simulation model of the 2DOF robot arm using the exact input - output feedback linearization method and pole placement with integrator method. The proposed control algorithm together with simulation model, which includes physical constraints and limits of the real model, was implemented into control structure and was verified in Matlab/Simulink program language. The resulting graph shows, that output of the model tracks step change of the reference trajectory and therefore can be considered, this approach is suitable for solution problem of control for 2DOF robot arm. The obtained knowledge from field of control algorithm design for MIMO systems using methods of nonlinear synthesis and also the proposed program module will be part of my dissertation work named Design of Effective Software Tools for Control and Analysis of Nonlinear Systems

ACKNOWLEDGMENT

This contribution is the results of the Vega project implementation: Dynamic hybrid architectures of the multiagent network control systems (No. 1/0286/11), supported by the Scientific Grant Agency of Slovak Republic (50%) and project implementation Development of the Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (50%).

REFERENCES

- [1] M. Razím and J. Štecha, "Nonlinear systems", ČVUT, 1997, (in Czech language).
- [2] G. Liu and A.A. Goldenberg, "Robust Control of Robot Manipulators Incorporating Motor Dynamics," Proc. of IEEE RSJ International Workshop on Intelligent Robots and Systems, Japan, 1993, pp. 68-74.
- [3] S. A. Mazhari and S. Kumar, "PUMA 560 Optimal Trajectory Control using Genetic Algorithm, Simulated Annealing and Generalized Pattern Search Techniques," International Journal of Electrical, Computer and Systems Engineering, Vol. 2, No. 1, 2008, pp. 71-80.
- [4] R. Palm, D. Driankov and H. Hellendoorn, "Model Based Fuzzy Control", Springer - Verlag Berlin Heidelberg, 1997
- [5] S. Senini and P. Wolfs, "Pole Placement Design of Integral Controller for Shunt Hybrid Active Filter", Australasian Universities Power Engineering Conference - AUPEC 2002, 29 sept.- 02 oct. 2002, pp. 160 - 165, Melbourne, Australia

Object Categorization using system Multi-Agent Serving System and robot Nao

¹Peter SMOLÁR (3rd year), ²Mária VIRČÍKOVÁ (2nd year)
Supervisor: ³Peter SINČÁK

^{1,2,3}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹peter.smolar@tuke.sk, ²maria.vircikova@tuke.sk, ³peter.sincak@tuke.sk

Abstract— In this paper, we present our work in categorization of real objects using camera and motion capability of humanoid robot. This system is built on Multi-Agent Serving System (MASS) created by PhD. student Tomáš Reiff in programming language Csharp. As a classifier of measured features is chosen modified neural network MF ARTMAP. Neural network MF ARTMAP in addition computes a membership value of every input sample for every output class and helps to make more human like interpretation of classification results. As a robotic platform is used robot Nao, humanoid robot made by the French company Aldebaran Robotics.

Keywords— object categorization, MF ARTMAP neural network, MASS system, robot Nao, fuzzy class.

I. INTRODUCTION

The question how will the robots stay and operate in human like environment is very exciting and a number of scientists solve these challenges to improve recognition and manipulation with an object in robot environment. It is mainly interesting for automation in industry, handling with explosive in military or for space robotic missions to the other planets.

In this paper, we want to present basic experiments of our system for robotic object categorization based on modular system Multi-Agent Serving System [4]. For purposes of categorization we used modified MF ARTMAP neural network. Basic version of this network gives us in the output values of membership to the each output class. Improved version is better adapted to the complexity of the input patterns. As a humanoid platform robot Nao was used.

Our goal is to show basic concept of the experiments with this robot for categorization of similar objects.

II. CATEGORIZATION

Categorization, as we understand it, is classification of objects (their attributes) in images to the more general classes based on their common (or different) attributes. For instance, categorization can be “generic” classes like human, car and town. Categorization task can be formulated as supervised or unsupervised learning (called also cluster analyze). In categorization basic problem is to handle large intra-class variability to one category and potentially small inter-class

differences between various categories [3].

Pinz in [3] characterize the object categorization as an issue of classification, learning, representation, localization, datasets, evaluation, system integration (Fig. 1). Classification deals with method of classification. We can decide for various types of classification methods (e.g. methods of machine learning, neural networks, Bayes decision rules, etc). Issue of learning learning deals with the way the acquisition of the knowledge. It can be supervised or unsupervised training. We have to decide for incremental or non-incremental learning. Representation concept defines how input features and output categories will be measured and represented. In some cases as 3D reconstruction or affine transformation, accurate localization is very important. It is necessary to locate each interest point of object for next processing. Entire proposal of categorization system depends on complexity of input images. This problem solves issue of datasets of samples. The important question is, how we will measure the results of categorization. We can choose various methods for it, for example recall and precision equation. Last issue deals with system integration of categorization method and robotic platform together for developing autonomous systems.

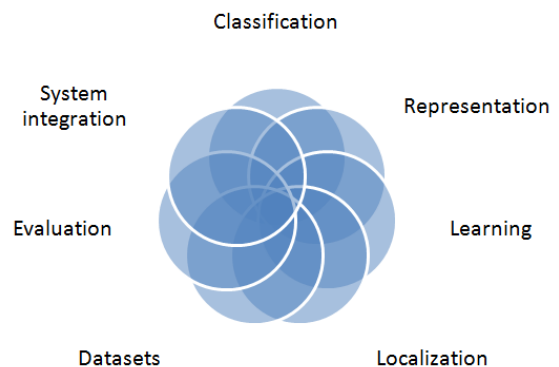


Fig. 1 Seven issues of object categorization.

In this work we propose approach for solving object categorization problem that is composed of two phases: Measurement of visual features and the classification. In measurement phase we propose and handle measurement of interest features of classified objects. In classification phase we propose and handle classification method for the categorization of input objects.

III. MULTI-AGENT SERVING SYSTEM (MASS)

Now, we briefly describe Multi-Agent Serving System (MASS) created by PhD. student Tomáš Reiff. It is a flexible and modular framework for working with various types of method and for connecting to various robots and units including various data sources. It has been designed as a client server system (Fig. 2). A client is running on the computer of logged user, a server is running on specific computer of better performance.

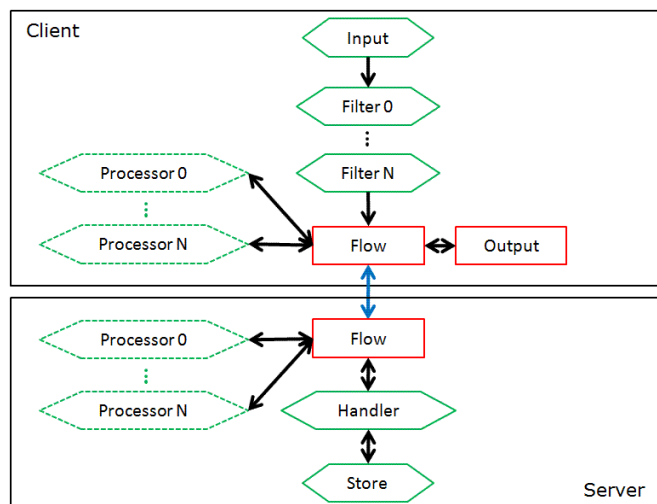


Fig. 2 The scheme of functional part of system MASS.

The user interface of client is created for selection of used plug-ins and for setting all basic parameters of the proposed system. The example of basic window is on Figure 3.

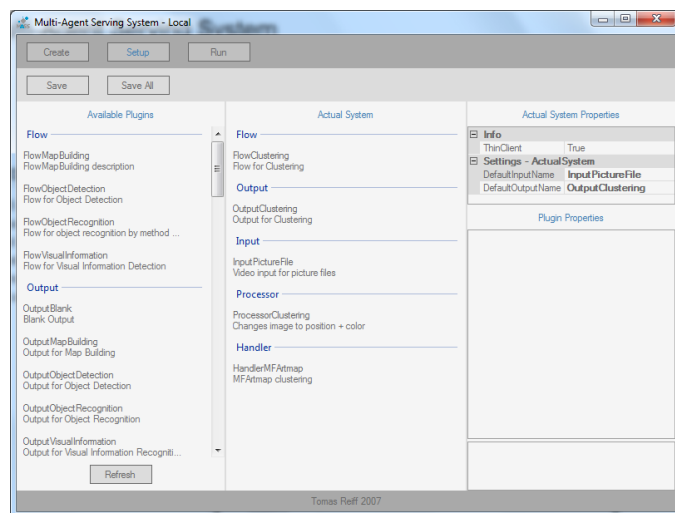


Fig. 3 Figure of system MASS with selected plug-ins.

IV. OBJECT CATEGORIZATION USING MODIFIED MF ARTMAP NEURAL NETWORK

Here, we will present two phases of object categorization process.

A. Measurement of visual features

In this step, we try to choose an appropriate method for

transforming an input image to the suitable form for classification. We decided to use detector&descriptor methods that are currently very popular and achieve good result in object recognition.

According to [3] it exists several factors of performance of such methods (e.g. repeatability, distinctiveness, locality, quantity, accuracy, efficiency, etc.). Under these conditions, we choose two methods: Scale invariant feature transform (SIFT) [2] and Speeded-Up Robust Features (SURF) [1]. In the Figure 4. descriptors of SIFT (left) and SURF (right) are showed.

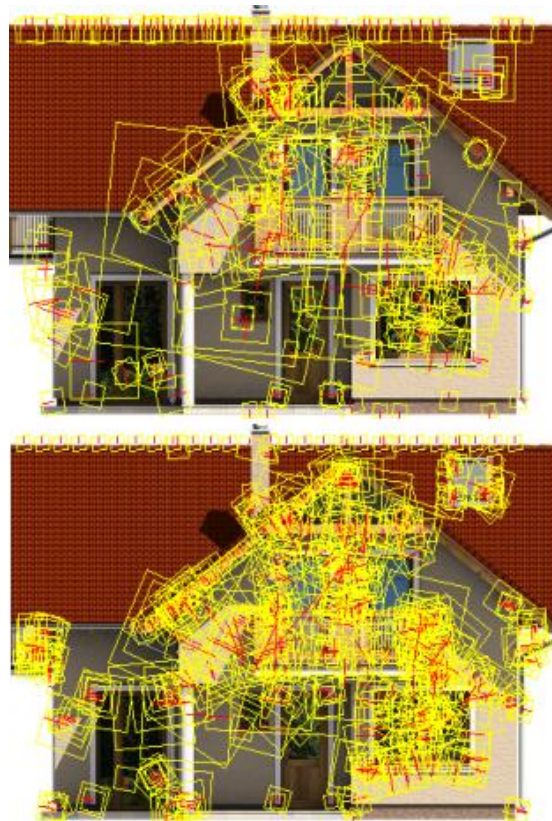


Fig. 4 Example of 285 SIFT descriptors (left) and 635 SURF descriptors (right).

B. Classification

For classification purposes we decide to use MF ARTMAP neural network based on ARTMAP family neural based classifiers with our modifications of clusters [5].

The great advantage of MF ARTMAP network is fuzzy output. It means that at the end, we know final class where input sample belong to and also membership value for each class. Other advantage is that, the class is represented by several clusters.

Modified MF ARTMAP network consists of three layers. First one is input layer, where features of input samples are mapped to the network. Compared to the classical MF ARTMAP our modification skips second comparison layer. Third layer is named as Recognition layer and this level represent clusters. Using our modification clusters are mathematically represented by function of multi-dimensional normal probability distribution. Last layer is "MAPFIELD", where output classes are mapped.

The basic concept of MF-ARTMAP neural network is on Figure 5.

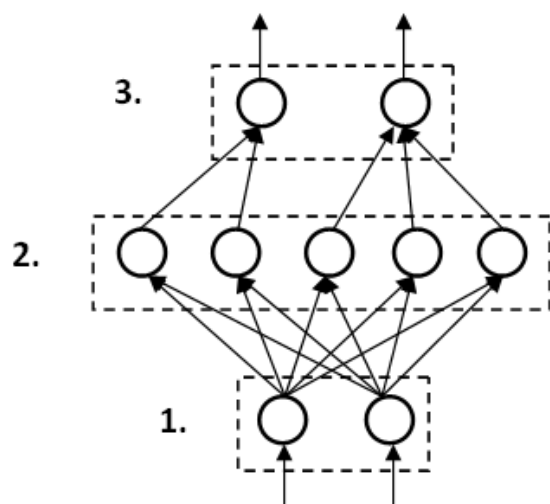


Fig. 5 Basic structure of modified MF-ARTMAP neural network described in text above. This network contains three layers: 1. Input layer, 2. Recognition layer, 3. MAPFIELD layer.

V. HUMANOID ROBOT NAO

In this section we briefly introduce a robot Nao of French company Aldebaran, which is used in our experiments. It is nowadays popular for education and research.

A. Hardware

It is medium-sized robot (58 cm) and it weighs 4.3 kilograms. Robot has two cameras video on its head.

It is medium-sized robot (58 cm) and it weighs 4.3 kilograms. Robot has two cameras (max. resolution: 640x480, visual angle: 34.8 degrees) and four microphones, distributed in its head. LEDs are distributed in five parts of body: inside of the tactile sensors (12), eyes (2x8), ears (2x10), torso (1), and feet (2x1). For connectivity is possible to use the Ethernet cable or the Wi-Fi (IEEE 802.11g).

B. Software

The NaoQi library contains special sub-modules like audio device, audio player, behavior manager, bonjour, face detection, file manager, frame manager, FSR, inertial, land mark detection, laser, launcher, LEDs, logger, logo detection, memory, motion, motion recorder, preferences, python bridge, sensors, sentinel, sonar, speech recognition, text to speech, video device, vision toolbox, DCM.

VI. PROPOSAL OF EXPERIMENTS

The structure of connection of our system is showed in Figure 6. In the computer is running system MASS (client and server in one PC). One server and two clients for two robots are established. Clients are connected to the robots using Wi-Fi connection. We use the front camera for video input from robots and this video is transmitted to the MASS (from robot to client and between client and server). After video processing and classification in the server, results are sent to the client.

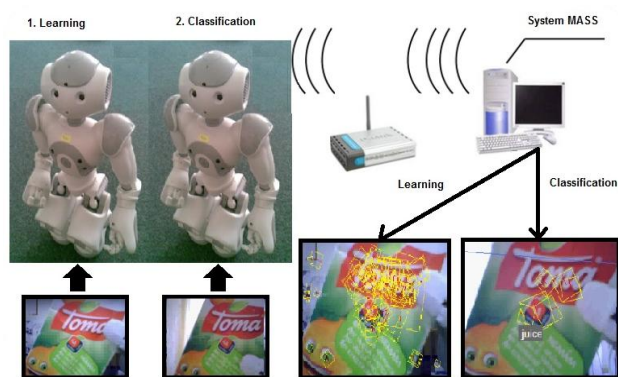


Fig. 6 Structure of our system with robot Nao for categorization of a TOMA juice box.

We used as a training sample juice box, beer box and box of fruit tee. In the Figure 7. the steps of experiment are showed with box of juice. Video of the experiment is available on the website: <http://www.ai-cit.sk/PeterSmolar>.

Proposed steps of experiment are:

1. Robot Nao #1 takes object before a camera.
2. Robot Nao #1 transmit video to the client of the system MASS. Server of system MASS learns an object as "juice".
3. Robot Nao #1 release this object from hands.
4. Robot Nao #2 takes the same or very similar object.
5. System MASS recognizes this object as a "juice".
6. Robot Nao no. 2 releases the object.

VII. CONCLUSION

In this article we tried to describe our approach of object categorization system. For this purpose, we used Multi-Agent Serving System (MASS), as a robotic platform was used robot Nao.

We performed a number of experiments and system was running with 81 % correct classification. Problems were mainly with Wi-Fi connection between MASS client and robot and with accuracy of motion. The object often fell down from hands of the robot.

In other work we will try improve reliability of the system.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

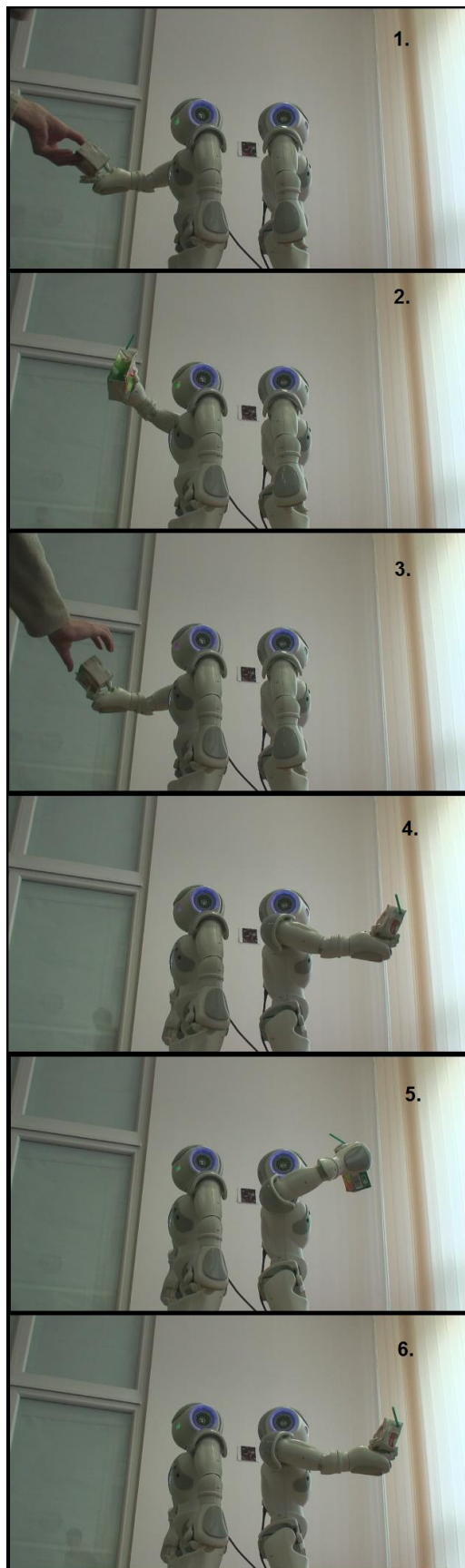


Fig. 7 The six steps of categorization experiment with two robots. First: Robot Nao #1 takes object before a camera. Second: Robot Nao #1 transmit video to the client of the system MASS. Server of system MASS learns an object as “juice”. Third: Robot Nao #1 release this object from hands. Fourth: Robot Nao #2 takes the same or very similar object. Fifth: System MASS recognizes this object as a “juice”. Sixth: Robot Nao no. 2 releases the object.

REFERENCES

- [1] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, Speeded-Up Robust Features, P. Simard, L. Bottou, P. Haffner, Y. LeCun, Boxlets: a fast convolution algorithm for signal processing and neural networks, in: NIPS, 1998.
- [2] D. G. Lowe, “Object recognition from local scale-invariant features,” in In Proc. International Conference on Computer Vision, pp. 1150–1157, 1999.
- [3] A. Pinz, Object Categorization, Foundations and Trends, in Computer Graphics and Vision, Volume 1 Issue 4, December 2005.
- [4] T. Reiff, Z. Fedor, M. Kuzma, Multi-agent serving system: an open source middleware for artificial intelligence and robotics. In: SCYR 2010 : 10th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovakia, FEI TU, 2010 S. 317-320. - ISBN 978-80-553-0423-6.
- [5] P. Smolár, J. Tuhársky, Improving the accuracy of classification and the generalization ability of an ARTMAP neural network / In: SCYR 2011 : 11th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice, May 17th, 2011, Herľany, Slovakia. - Košice : FEI TU, 2011 S. 291-293. - ISBN 978-80-553-0644-5.

On decomposition of High Level Petri Nets

¹Ivan PETKO (4th year),
Supervisor: ²Štefan HUDÁK

^{1,2}Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹ivan.petko@toryconsulting.sk

Abstract—The paper deals with semi-formal analysis of principles of decomposition of High Level Petri Nets with respect to the principles used for Low Level Petri Nets concentrating on the top level approach and forming the main ideas which hold for every method that may be used in the HLPN case. The process of decomposition and the issues concerning the difference between low level and high level are analysed. Decomposition is considered in accordance with the existing approaches as splitting relevant places and/or transitions.

Keywords—Decomposition of net notation, extension of decomposition principles, High Level Petri Nets, structural decomposition.

I. INTRODUCTION

Petri nets (PN) are a well-established formalism providing a simple and intuitive mechanism to describe and analyse discrete systems in a very comfortable way. In order to reflect growing requirements of real life systems the formalism is developed, examined and extended to this day improving its capability to model and analyse the systems (stochastic/time extensions [6], adapting external ideas such as object oriented perspective [9], algebraic approach [11], [14] and so on). These systems are getting more complex structure hand-in-hand with their evolution. This turns out to be a weakness of the whole formalism, since even small systems may require large nets hard to understand and analyse. The growing complexity requires more sophisticated solutions based on more efficient approaches such as de/composition e.g. [13], [2], [4] and more compact representations [5] of the systems.

During the research in the field of PN several useful extensions were proposed since PN were first introduced by C.A.Petri [7] in their basic form. The most useful one with respect to overcome the complexity weakness is a high-level concept (High Level Petri Nets - HLPN) providing more compact representations of systems compared to the low-level one. The "compactness" of HLPN is mainly given by their ability to distinguish between tokens – token individuals and whole behaviour is adapted to this end – HLPN use transition conditions and arc annotations making the concept close to the functional one during the execution of a net. Compared to low-level PN, HLPN are much harder to analyse, since in the low-level case we consider only one type of tokens and markings are represented by simple vectors thus in most cases the standard algebra and some basic principles of the graph theory suffice to analyse the low-level nets. In the high-level case one has to consider arc annotations, transition conditions and various types of tokens resulting in multiset markings. In order to analyse a high-level PN the two basic approaches

may be used – unfolding of the net [8], [3] into a low-level one, i.e. reducing the problem to the low-level case and exploit the analysis methods proposed for low-level nets or treat the net as a compact representation (HLPN) which is much harder (maybe not always possible) case. Another way, similarly as for large low-level nets, is taking advantages of a de/compositional approach to make the analysis easier.

In the following we focus on decomposition of the chosen class of HLPN defined in the international standard [5]. At first the HLPN class definition is introduced. Subsequently the process of HLPN decomposition is considered with respect to corollaries rising up from the high level case in a semi formal way. We concentrate on decomposition from a general point of view as splitting relevant places and/or transitions.

II. HLPN DEFINITION

In order to investigate decomposition of HLPN we focus on the HLPN standard [5]. The authors of the standard claim that it covers the ideas forming basic HLPN classes namely Pr/T nets [12], colored nets [1] and algebraic nets [10]. There are some preliminaries we leave out in this paper such as multiset or formal term definitions. For more detailed information we refer to [5]. The standard includes two main definitions – HLPN and HLPN graphs. Since decomposition is more illustrative in the case of HLPN graphs, we consider them a base for our treatment and refer to this class as $HLPN(G)$.

Definition 1: HLPN graph ($HLPNG$) is a structure

$$HLPNG = (NG, Sig, V, H, Type, AN, m_0),$$

where

$NG = (P, T, F)$ is a net graph with

P – set of places

T – set of transitions

$F \subseteq (P \times T) \cup (T \times P)$ – set of directed arcs referred to as flow relation

$Sig = (S, O)$ is a many-sorted Boolean signature with the set of sorts $S = \{Boolean\}$

and boolean operations $O = \{<, \leq, \vee, =, \dots\}$

V – S -indexed set of variables, $V \cap O = \emptyset$

$H = (S_H, O_H)$ – many-sorted algebra for Sig defining its meaning

$Type : P \rightarrow S_H$ – mapping assigning types (sorts) to places

$AN = (A, TC)$ – net annotation with

$A : F \rightarrow Term(O \cup V)$ – mapping assigning terms to arcs. The result of term evaluation is a multiset over the types of associated

places, i.e. $\forall((p, t), (t', p) \in F) \forall \alpha :$
 $Val_\alpha(A(p, t)), Val_\alpha(A(t', p)) \in$
 $Bag(Type(p))$, where $Term(O \cup V)$ is a set
 of terms over variables and operations,
 α is an assigning of token values to variables,
 $Val_\alpha(term)$ is term evaluation and $Bag(B)$
 is a set of multisets over B
 $TC : T \rightarrow Term(O \cup V)_{Bool}$ is a mapping
 which assigns boolean expressions
 to transitions
 $m_0 : P \rightarrow \bigcup_{p \in P} Bag(Type(p))$ is initial marking

Definition 2: Marking of $HLPNG$ =
 $(NG, Sig, V, H, Type, AN, m_0)$ is a mapping

$$m : P \rightarrow \bigcup_{p \in P} Bag(Type(p)),$$

such that $\forall p \in P : m(p) \in Bag(Type(p))$.

Based on the definition of PN as bipartite graphs we refer to the subsets of transitions and places forming the interface of decomposition given for $HLPN \in HLPN(G)$ as $I_d = T_d \cup P_d$, where $T_d \subseteq T$, $P_d \subseteq P$ in accordance with general approaches to decomposition.

III. HLPN DECOMPOSITION

As outlined above the main added value of HLPN compared to the low level concept is using of arc expressions and transition conditions what follows immediately from their capability to differ individual tokens mathematically represented as multisets [5].

Formally decomposition may be considered a function on $HLPN(G)$ domain

$$\delta_C : HLPN(G) \times I_d \rightarrow HLPN(G)^2,$$

where $I_d \subseteq P \cup T$ - interface of decomposition,
 $C \in \{P, T, PT\}$.

Index C determines the type of decomposition and thus I_d as well, particular types are described below. It is obvious that there is at least one minimal component $N_\varepsilon \in HLPN(G)$ irreducible in terms of such decomposition, i.e. $\delta(N_\varepsilon) = N_\varepsilon$. In other words there may be irreducible nets in terms of the proper decomposition used dependent on conditions to be kept. We remind that decomposition to n components is achieved by $n-1$ applications of δ to its particular results (components) what justifies the definition of δ as a function transforming a component to two subcomponents.

Definition 3: Let $N = (NG, Sig, V, H, Type, AN, m_0) \in HLPN(G)$. Then $(N_1, N_2) = \delta_C(N)$, $N_i = (NG_i, Sig_i, V_i, H_i, Type_i, AN_i, m_{0i})$, $i = \{1, 2\}$ are given as follows:

$$\begin{aligned} (P_1 \cap P_2) \setminus I_d &= (T_1 \cap T_2) \setminus I_d = \emptyset \\ \forall p \in P_d : p &\in P_1 \wedge p \in P_2, \forall t \in T_d : t \in T_1 \wedge t \in T_2 \\ F_1 \cap F_2 &= \emptyset \text{ for } \delta_P, \delta_T, F_1 \cap F_2 = \{(p, t), (t, p) \in F | p \in P_d, t \in T_d\} \text{ for } \delta_{PT} \\ F_1 \cup F_2 &= F \\ Sig_1 &= Sig_2 = Sig \\ V_1 &= V_2 = V \\ H_1 &= H_2 = H \\ Type_1 &= Type_2 = Type \\ \forall f \in F_i : A_i(f) &= A(f) \end{aligned}$$

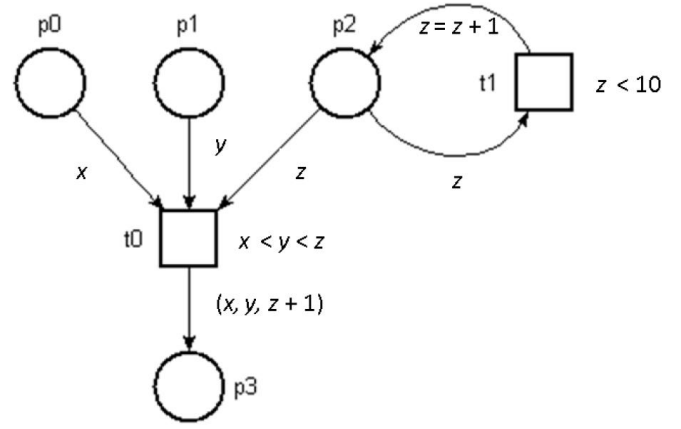


Fig. 1. Fragment of HLPN

$$\forall t \in T_i : TC_i(t) = TC(t)$$

$m_{0i} = m_0|_{P_i}$, i.e. restriction of m_0 to places from P_i , $m_{0i} \in Bag(P)$, $m_{0i} \in Bag(P_i)$.

There may be some redundant elements in Sig_i , V_i , H_i , $Type_i$ in particular subnets, for instance not all variables are contained in all and arbitrarily small subnets larger or equal to some N_ε but in general the appearance of variables in more subnets can not be excluded. It is obvious that this may hold for signature operators, sets of values for types and type functions analogously.

In order to consider net annotation decomposition and with respect to the ideas described below we introduce the notion of incidence places for transitions.

Definition 4: The set of incidence places for transition $t \in T$ in $HLPN(G)$ $N = (NG, Sig, V, H, Type, AN, m_0)$ is a set

$$\begin{aligned} P_t &= \{p \in \bullet t | \exists v \in vars(TC(t)) : p \in src(v) \vee \exists f \in \\ &(((\{t\} \times P) \cup (P \times \{t\})) \cap F) : \exists v \in vars(A(f)) : p \in src(v)\}, \end{aligned}$$

where

$vars(term)$ – set of variables contained in $term$ including global variables

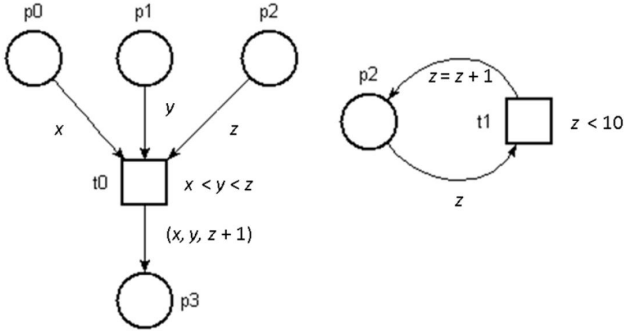
$src(v) = \{p \in P | v \in TC(t) \wedge t \in p \bullet \vee v \in A(f) \wedge f \in (p \times T) \cup (p \bullet \times P)\}$ – set of source places for variable v , i.e. the places holding the tokens which may be assigned to v .

The set of incidence places for a subset of transitions $T_1 \subseteq T$ is given by $P_{T_1} = \bigcup_{t \in T_1} P_t$.

As outlined above Petri nets as bipartite graphs may be decomposed only in the following ways:

- 1) creating subnets by duplicating chosen contact places, i.e. through $I_d = P_d \subset P$, in terms of convention introduced in [13] we denote this approach P decomposition, δ_P
- 2) creating subnets by duplicating chosen contact transitions, i.e. through $I_d = T_d \subset T$, in terms of convention introduced in [13] we denote this approach T decomposition, δ_T
- 3) combined approach, i.e. $I_d = P_d \cup T_d$, $P_d \neq \emptyset$, $T_d \neq \emptyset$, denoted PT decomposition [13], δ_{PT}

In the process of HLPN decomposition in contrast to the low level one when only the structural approach is needed the following has to be taken into account:


 Fig. 2. P decomposition $P_d = \{p2\}$, $T_d = \emptyset$

- 1) structural decomposition
- 2) decomposition of net annotation - depends on the structural decomposition which may affect some arc expressions and/or transition conditions

A. Structural HLPN decomposition and its corollaries

a) *P decomposition:* Some of the results obtained in the field of decomposition for low level nets may be applied with no need to extend, actually all the approaches decomposing a net through $I_d = P_d \subset P$ and $T_d = \emptyset$. Moreover structural P decomposition is quite straightforward – there is no need to perform any additional decomposition of the annotations since P decomposition does not affect it at all, therefore we state the following proposition.

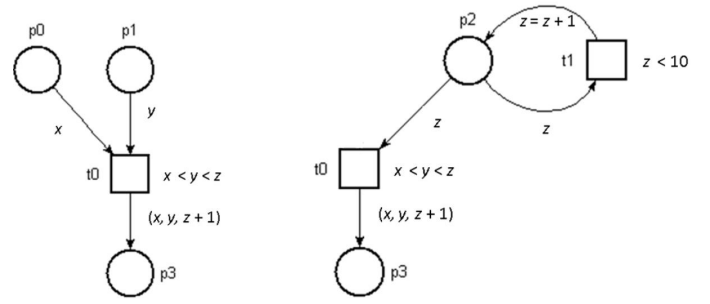
Remark 1: Let $N \in HLPN(G)$. Then each low level approach to decomposition may be applied without changes if $P_d \subseteq P$ and $T_d = \emptyset$ and no global variables are used in N and the result of decomposition is the same as in the low level case.

This follows immediately from the P decomposition principle, i.e. duplicating places $p \in P_d$ in relevant subnets (Fig. 2) what ensures that no variables may occur free in any subnet thus all variables may be assigned in the particular subnets. It is obvious from the $HLPN(G)$ definition that $p \in P_d$ are the sources for assigning the variables occurring in the incidence transitions and arcs, more precisely $\forall v \in V$ such that $\exists t \in T : v \in vars(TC(t)) \vee \exists f \in F : v \in vars(A(f))$ we have $\exists p \in P : p \in src(v)$, where $v \in vars(TC(t)) \wedge t \in p \bullet \vee v \in vars(A(f)) \wedge f \in (p \bullet \times T) \cup (p \times P)$, provided that each variable is local, i.e. it may be assigned only in the scope of arcs between p and $p \bullet$, transitions $t \in p \bullet$ and arcs between $p \bullet$ and P .

Since $p \in P_d$ are source places for variables appearing in the incidence arc expressions and transition conditions it is ensured that no variable is unassigned after decomposition regarded as duplicating of the places $p \in P_d$ what directly leads to the fact that the annotation (arc expressions and transition conditions) is not influenced.

Notice that if N before decomposition contains global variables these would be unassigned in the relevant subnets what justifies the condition the net annotation is to be without global variables.

b) *T and PT decomposition:* We may notice that in some cases the net annotation is affected, i.e. free variables may appear when performing HLPN decomposition, specifically if $T_d \subset T$, that is T and PT decomposition. The situation is depicted in the Fig. 3 with free variables appearing – z in


 Fig. 3. PT decomposition $P_d = \{p3\}$, $T_d = \{t0\}$

the first subnet, x and y in the second one. Simulation of these subnets separately is not possible without changing the affected expressions since their behaviour is given by variable assigning and free variables remain non-assigned.

Modification of subnet annotations may change the originally intended behaviour since in order to eliminate free variables the contribution of adjacent subnets (subnets with common contact places/transitions after decomposition) is removed as a corollary of the fact that source places for the free variables are not included.

With respect to the sketched problem appearing in T and PT decomposition we may proceed twofold:

- 1) transformation of T/PT to P decomposition by adding the incidence places P_t for each $t \in T$ such that free variables may appear after T/PT decomposition, i.e. all source places for all variables used in $TC(t)$ and incidence arcs between t , $\bullet t$ and t , $t \bullet$
- 2) perform T/PT decomposition with necessary decomposition of net annotation for all expressions incorporating free variables in the particular subnets

In the first case it is trivially ensured that no free variables may occur since P decomposition guarantees that all relevant places meant as source places are always attended to each transition requiring variable assigning from the places. But transformation of T/PT to P decomposition by adding the incidence places P_t for each $t \in T$ such that free variables may occur after T/PT decomposition has an important corollary:

Corollary 1: Some of the approaches to decomposition used for low level nets can not be applied directly – the resulting subnets may differ in the HLPN case. For instance decomposition based on index function [2] produces a set of so called S-nets ($\forall t \in T_{S-net} : (|\bullet t| \leq 1 \wedge |t \bullet| \leq 1)$). In the high level case the same effect is not possible because of including incidence places P_t and the algorithm for index decomposition would have to be changed itself.

c) *Conditions for decomposition of net annotation:* Decomposition of net annotation is necessary in the case of T or PT decomposition or any other algorithm leading to T or PT decomposition no matter if performed manually or automatically. As indicated above decomposition of net annotation is applied to affected terms, i.e. the terms containing free variables as a consequence of omitting some source places for the variables in particular subnets. It is obvious that decomposition under consideration means term modifying in such a way that eliminates free variables without affecting the subnet execution semantics and the sequel of the term. It follows immediately that it is possible to introduce conditions for decomposition of net annotation.

Let $N_1, N_2 \in HLPN(G)$ be two subnets of net N being decomposed, i.e. $\delta_{T/PT}(N) = (N_1, N_2)$. Then the transition condition $TC(t)$ of a transition t has to be decomposed in terms of omitting free variables \iff

$$\exists p \in P_j \setminus P_i :$$

$$((p, t) \in F \setminus F_i \vee (t, p) \in F \setminus F_i) \wedge \exists v \in vars(TC(t)) : p \in src(v) \\ i, j \in \{1, 2\}, i \neq j$$

Similarly the arc expression $A(f)$ of an arc f has to be decomposed \iff

$$\exists t \in T_i : \exists p \in P_j \setminus P_i :$$

$$((p, t) \in F \setminus F_i \vee (t, p) \in F \setminus F_i) \wedge \exists f \in \{(p, t), (t, p)\} \in F \setminus F_i : \\ \exists v \in vars(A(f)) : p \in src(v), i, j \in \{1, 2\}, i \neq j$$

Note that v denotes a free variable in the statements above (initially a local variable in a part of net which is separated after decomposition from another or a global variable) thus the statements may be considered conditions for free variables generation. In other words the statements generate sets of free variables after decomposition.

IV. CONCLUSION

Decomposition of high level Petri nets by adopting the low level principles and investigating of their extensions to the high level case is analysed from a general point of view in the paper. Instead of providing an algorithm different ways of decomposition are examined – decomposing a net through a set of common places, common transitions or both. Decomposition through places is fully safe in terms of avoiding modification of the net annotation (transition conditions and arc expressions) and the low level principles may be applied directly since no variable becomes free in any subnet (common places are duplicated). The other manners may result in modifying the net annotation if the source places for the variables used in the subnet annotation are included only in the other subnet(s). In this case free variables may occur and the subnet notation must be modified in order to eliminate the free variables and the relations between subnets kept in the transition conditions and arc annotations may be lost. In order to avoid such a modification decomposition including the so called incidence places is proposed. Another ways of decomposition are investigated as well and some issues concerning the process are sketched including stating the conditions for decomposition of net annotation.

REFERENCES

- [1] Jensen, K.: Coloured Petri Nets: A high-level Language for System Design and Analysis, LNCS vol. 483, Springer Verlag, 1990
- [2] Zeng, Q., Hu, X., Zhu, J., Duan, H.: A polynomial-time decomposition algorithm for a Petri net based on indexes of places, Journal of Applied Sciences 8(24): 4668-4673, 2008, ISSN 1812-5654, 2008, Asian Network for Scientific Information
- [3] Khomenko, V., Koutny, M.: Branching Processes of High-Level Petri Nets, Proc. of TACAS'2003, LNCS 2619 (2003) 458–472.
- [4] Zaitsev, D. A.: Decomposition of Petri nets, Cybernetics and Systems Analysis, no. 5, pp. 131–140, 2004
- [5] ISO/IEC 15909-1:2004, High-level Petri Nets - Concepts, Definitions and Graphical Notation, http://www.iso.org/iso/catalogue_detail.htm?csnumber=38225
- [6] Gianfranco, B.: Introduction to stochastic Petri nets, Springer Lectures On Formal Methods And Performance Analysis, Lectures on formal methods and performance analysis: first EEF/Euro summer school on trends in computer science, pp. 84–155, 2002, ISBN:3-540-42479-2
- [7] Petri, C.A.: Kommunikation mit Automaten, Bonn: Institut für Instrumentelle Mathematik, Schriften des IIM Nr. 2, 1962, Second Edition., New York: Griffiss Air Force Base, Technical Report RADC-TR-65-377, Vol.1, 1966, pp: Suppl. 1, English translation
- [8] Khomenko, V.: Model Checking Based on Prefixes of Petri Net Unfoldings, PhD Thesis, School of Computing Science, University of Newcastle upon Tyne, 2003
- [9] Battiston, E., DeCindio, F., Mauri, G.: OBJSA nets: a class of high-level nets having objects as domains, Springer Lecture Notes In Computer Science, Advances in Petri Nets 1988, pp. 20-43, ISBN:3-540-50580-6, 1988
- [10] Vautherin, J.: Parallel systems specifications with Coloured Petri nets and algebraic specifications, European Workshop on Applications and Theory of Petri Nets, 1986, pp. 293–308
- [11] Best, E., Devillers, R., Koutny, M.: Petri net algebra, 2001, XI, 378 p., Hardcover, ISBN: 978-3-540-67398-9
- [12] Genrich, H.J.: Predicate/transition nets, Advances in Petri nets 1986, part I on Petri nets: central models and their properties, pp. 207 — 247, ISBN:0-387-17905-4, Springer-Verlag, 1987
- [13] Hudák, Š.: Reachability analysis of systems based on Petri nets, Košice, TU, 1999, ISBN: 80-88964-07-5
- [14] Best, E., Devillers, R., Hall, J.: The petri box calculus: A new causal algebra with multilabel communication, in Advances in Petri Nets, Lecture Notes in Computer Science, Vol. 609, pp. 21–69, Springer-Verlag, 1992

On Kronecker product of graphs

¹Matúš VALO (1st year)

Supervisor: ²Marián KLEŠČ

^{1,2}Dept. of Mathematics and Theoretical Informatics, FEI TU of Košice, Slovak Republic

¹matus.valo@tuke.sk, ²marian.klesc@tuke.sk

Abstract—The crossing number $cr(G)$ of the graph G is the minimum number of edge crossings in any drawing of G . Regarding to crossing number, Cartesian product of two graphs has received the greatest attention among products of graphs. However some interesting results have appeared recently. This paper gives a survey of results known about crossing number of Kronecker product of two graphs and adds another new results on the crossing numbers of this product.

Keywords—Crossing number, Kronecker product

I. INTRODUCTION

A finite, undirected, simple graph $G = (V, E)$ is an ordered pair consisting of finite set V of vertices and finite set of edges $E \subseteq \{\{u, v\} | u \in V, v \in V, u \neq v\}$. The vertex and edge set of the graph G we will denote by $V(G)$ and $E(G)$, respectively. We will abuse terminology and call finite, undirected, simple graph as *graph*. The graphs are also connected unless it is not explicitly stated. The edge we will denote by uv instead of $\{u, v\}$. A path and cycle with n vertices we will denote P_n and C_n , respectively.

Graphs are usually represented in a plane by visualising each vertex as a single point and each edge as a curve connecting points which represent end vertices of given edge. We will consider drawings with following properties:

- 1) adjacent edges never cross,
- 2) two non-adjacent edges cross at most once,
- 3) no edge crosses itself and
- 4) no more than two edges cross at a point of the plane.

The *crossing number* of a graph G is denoted by $cr(G)$ and is defined as the smallest number of edge crossings in any drawing of G .

Finding a crossing number is useful in many areas. The most prominent area is VLSI technology. The lower bound on the chip area is determined by crossing number and by number of vertices of the graph [1], [2]. It plays an important role in various fields of discrete/computational geometry [3]. Crossing number is also parameter yielding the deviation of the graph from being planar. The crossing number influences significantly readability and therefore it is the most important parameter when considering aesthetics of a graph. It is used mostly in automated visualisation of graphs.

However finding a crossing number of a given graph is very hard. It was proved by Garey and Johnson [4] that the crossing number of a graph is NP-complete and it remains NP-complete even for cubic graphs [5].

Very special type of drawing of a graph is an orthogonal drawing, which allows only $\pi/2$ or π angle between two adjacent edges. This type of drawing has several application.

The basic abstraction underlying a software system running on a set of distinct machines consists of a set of finite transition systems. An orthogonal drawing gives an intuitive representation of such systems [6].

Computer hardware and microchips are designed using CAD tools, which must create a layout of the logic gates and their interconnections on circuit boards. The layout themselves corresponds to a grid drawing in which all vertices and bends of the edges have integer coordinates [7]. An orthogonal drawing is similar to a grid drawing.

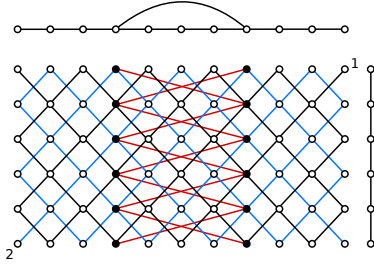
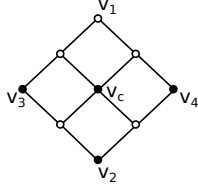
Entity-relationship diagrams are one of the common methods to aid structuring large volumes of data by defining attributes and relations between the data [6]. Both entities and their attributes are modeled as nodes of a graph. Edges express that given attribute is an attribute of given entity by linking an attribute to an entity. The resulting graph is presented by means of an orthogonal drawing.

II. KRONECKER PRODUCT

Let G and H are graphs. The *Kronecker product* $G \times H$ is a graph with vertex set $V(G \times H) = V(G) \times V(H)$ and edge set $E(G \times H) = \{\{(x_g, x_h), (y_g, y_h)\} | x_g y_g \in E(G) \text{ and } x_h y_h \in E(H)\}$. This product is also known as direct product, tensor product, cardinal product, cross product and graph conjunction.

The Kronecker product $C_m \times C_n$ is an important class, which has many useful properties. The most prominent property is that such graph is a four-regular graph, even further it contains subgraphs isomorphic to a grid. It enables orthogonal drawing of such graph. Jha and Devisetty formulated algorithms for orthogonal embedding $C_m \times C_n$ [8]. The *Cartesian product* $G \square H$ of the graphs G and H is a graph with the vertex set $V(G \square H) = V(G) \times V(H)$ and the edge set $E(G \square H) = \{\{(x_g, x_h), (y_g, y_h)\} | x_g y_g \in E(G) \text{ and } x_h = y_h, \text{ or } x_h y_h \in E(H) \text{ and } x_g = y_g\}$. If m and n are both odd, then $C_m \times C_n$, also known as diagonal mesh [9], [10], [11], has a lower diameter, higher independence number and higher odd girth relative to $C_m \square C_n$, also known as toroidal mesh. Pearlmutter showed that a diagonal mesh is isomorphic to a twisted toroidal mesh [12] and that a twisted toroidal topology was earlier used as the routing network of the FAIM-1 parallel computer [13].

Kronecker product has also further interesting properties. P. M. Weichsel proved assertion that $G \times H$ is connected if and only if either G or H contains an odd cycle. If neither of them contains an odd cycle then $G \times H$ contains exactly two connected components [14]. Let a 1-contraction be the removal each vertex of degree 1 from G . M. Farzan and D. A. Waller proved following theorem [15]:


 Fig. 1. A drawing of graph $A_{3,3}^5 \times P_6$.

 Fig. 2. Envelope $\text{env}(v_c)$

Theorem 1: Let G_1 and G_2 be connected graphs with more than four vertices. Then $G_1 \times G_2$ is planar if and only if either

- 1) one of the graphs is a path and the other one is 1-contractible to a path or circuit, or
- 2) one of them is a circuit and the other is 1-contractible to a path.

Let us define a *decomposition* \mathcal{D} of a graph G as a set of subgraphs G_1, \dots, G_r which constitute a partition of the edge set of G . A *factorization* \mathcal{F} of a graph G consists of spanning subgraphs F_1, \dots, F_r which constitute a partition of the edge set of G . The spanning subgraphs F_1, \dots, F_r are called *factors* of G . A biparted graph on an even number of vertices is said to admit of a *bi-pancyclic ordering* if there is an ordering $v_0, v_1, \dots, v_{2r-1}$ of its vertices such that $\langle v_0, v_1, \dots, v_{2k-1} \rangle$ contains a spanning cycle for all $k \in \{2, \dots, r\}$. P. K. Jha studied decomposition, factorisation and bi-pancyclic ordering of the Kronecker product of cycles and paths. His main result is the following [16]:

- 1) If m is even and $n \equiv 0 \pmod{4}$, then one component of $P_{m+1} \times P_{n+1}$, and each component of each of $C_m \times P_{n+1}$, $P_{m+1} \times C_n$ and $C_m \times C_n$ are edge decomposable into cycles of uniform length rs , where r and s are suitable divisors of m and n , respectively.
- 2) If m and n are both even, then each component of each of $C_m \times P_{n+1}$, $P_{m+1} \times C_n$ and $C_m \times C_n$ is edge-decomposable into cycles of uniform length ms , where s is suitable divisor of n ,
- 3) $C_{2i+1} \times C_{2j+1}$ is factorizable into shortest odd cycles,
- 4) each component $C_{4i} \times C_{4j}$ is factorizable into four-cycles and
- 5) each component of $C_m \times C_{4j}$ admits of a bi-pancyclic ordering.

The crossing number of $C_m \times C_n$ is so far unknown, but Jha and Devisetty proved its lower bound [8]:

Theorem 2: $\text{cr}(C_m \times C_n)$ is greater than or equal to

- 1) $(0.8 - \epsilon)mn$ m, n is odd and equal, $m \geq n_0$
- 2) $1/8(0.8 - \epsilon)mn$ m is odd, n is even, $n \geq 6$,
 $\min\{m, n/2\} \geq n_0$
- 3) $1/8(0.8 - \epsilon)m(n - 1)$ m, n is odd, $m < n$, $n \geq 7$,
 $\min\{m, (n - 1)/2\} \geq n_0$

where, in each case, $\epsilon > 0$ and n_0 is a sufficiently large

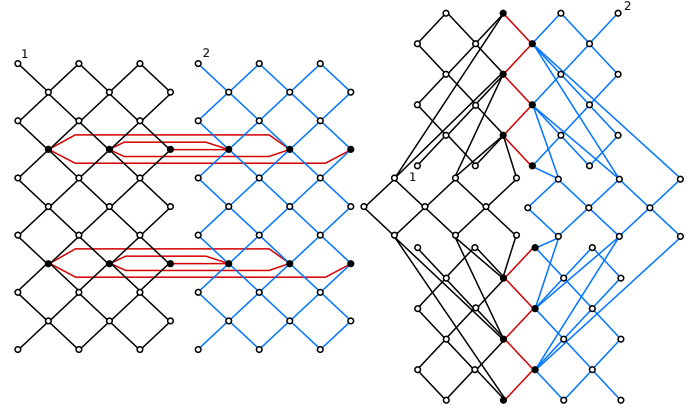


Fig. 3. Two possible separations of a grid.

integer depending only on ϵ .

and upper bound too:

Theorem 3: Let m be odd. Then $\text{cr}(C_m \times C_n)$ is less than or equal to

- | | |
|-----------------------------------|------------------------------------|
| 1) $(m - 2)n$ | $n = m$ |
| 2) $(m - 1)n$ | $n = km, k$ is even,
$k \geq 2$ |
| 3) $(m - 1)n + m$ | $n = km, k$ is odd,
$k \geq 3$ |
| 4) $(2m - n)(n - 3) + n(n/2 - 1)$ | n is even, $m \geq n/2$ |
| 5) $(n - 2m)(2m - 3) + 2m(m - 1)$ | n is even, $n/2 > m$ |
| 6) $2mn - (m^2 + 3m)$ | n is odd,
$m < n < 3m$ |
| 7) $(2mn + 6m) - (2m^2 + 3n)$ | n is odd, $n \geq 3m$ |

Let P_s be a path on s vertices, where $s > 5$, with vertex set $V(P_s) = \{v_1, v_2, \dots, v_m, \dots, v_{m+k}, \dots, v_s\}$. Then $A_{m,n}^k$, $n = s - m - k$ is a graph created from P_s by adding a new edge (v_{m+1}, v_{m+k}) . See Fig. 1 showing a graph $A_{3,3}^5 \times P_6$. A subgraph isomorphic to P_s of this graph we will denote by $P(A_{m,n}^k)$.

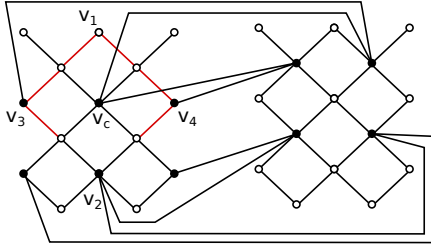
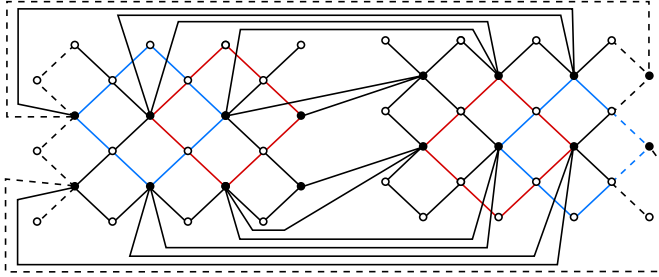
Botureau asserts that for two graphs G and H such that G is subgraph H , it holds that for every graph K , $G \times K$ is a subgraph of $H \times K$ [17]. This property of the Kronecker product $A_{m,n}^k \times P_s$ implies that this graph contains two disjoint subgraphs isomorphic to subgraphs of a grid. Furthermore, there exist also another four subgraphs of $A_{m,n}^k \times P_s$ isomorphic to subgraphs of a grid. By separation of this grids we can gain a drawing of the graph with crossings only between edges connecting this grids with edges of the grids. Fig. 3 illustrates this drawing of the graph $A_{3,3}^5 \times P_6$. Such drawing has a great chance to have minimum number of crossings. We will prove this property for every graph $A_{m,n}^k \times P_s$, $m, n \leq 2$, $k \leq 3$.

Let E_c be a set of all edges connecting two grids of the graph $A_{m,n}^k \times P_s$. We will call such edges *connection edges*. The vertices incident with connection edges we will call *connection vertices* and the set of all such vertices we will denote by V_c .

There exists a class of planar graphs $A_{m,n}^k \times P_s$ with k odd. The following theorem results from the Theorem 1:

Theorem 4: $A_{m,n}^k \times P_s$ is planar if and only if $m, n \leq 1$ or $s \leq 4$.

Let G be one of the grids of the graph $A_{m,n}^k \times P_s$ and $v_c \in V_c$ be one of the connection vertices of the grid G such that it is not incident with outer face in some planar drawing of G . Then there exists subgraph G^* isomorphic to the graph


 Fig. 4. A drawing of graph $A_{2,1}^3 \times P_5$

 Fig. 5. A drawing of graph G_7 and G_8

on the Fig. 2 containing the vertex v_c , $\deg_{G^*}(v_c) = 4$. Such subgraph we will call an *envelope* of the vertex v_c and denote $\text{env}(v_c)$. The vertex v_c is a *central vertex* of the graph $\text{env}(v_c)$.

Lemma 1: $\text{cr}(A_{2,1}^3 \times P_5) = 2$

Proof: From Fig. 4 it is clear that $\text{cr}(A_{2,1}^3 \times P_5) \leq 2$.

Let us designate $A_{2,1}^3 \times P_5$ as G and let v_c be connection vertex having envelope $\text{env}(v_c)$. It can be easily seen that $\text{cr}(G) \geq 1$ since it contains subgraph homeomorphic to the graph $K_{3,3}$. Now by way of contradiction suppose that $\text{cr}(G) = 1$. It can be easily verified that removing any black edge e on the Fig. 4 maintains non-planarity of the graph $G - e$, since the graph $G - e$ still contains a subgraph homeomorphic to $K_{3,3}$. It implies that only red edges can cross each other. Let G^p be a graph created from the graph G by crossing two "allowed" edges and replacing the crossing with a new vertex \tilde{v} . Such graph must be planar, since we suppose that $\text{cr}(G) = 1$. It can be easily verified that every graph G^p created by crossing any of two "allowed" edges contains subgraph homeomorphic to $K_{3,3}$, contradicting the hypothesis $\text{cr}(G) = 1$. ■

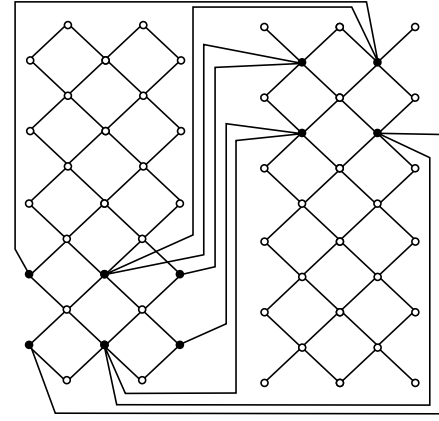
Corollary 1: $\text{cr}(A_{n,1}^3 \times P_5) = 2$, $n \geq 2$

Proof: Crossing number of graph $A_{n,1}^3 \times P_5$ is at most 2 provided by Fig. 6, which shows an example of the embedding of the graph $A_{7,1}^3 \times P_5$. Since the graph $A_{n,1}^3 \times P_5$ contains $A_{2,1}^3 \times P_5$ as a subgraph, it implies that $\text{cr}(A_{n,1}^3 \times P_5) \geq 2$. ■

Theorem 5: $\text{cr}(A_{2,1}^3 \times P_s) = 2(s - 4)$, $s \geq 5$

Proof: We will present only sketch of the proof. Let G and G' denote graphs $A_{2,1}^3 \times P_s$ and $P(A_{2,1}^3) \times P_s$, respectively. It can be easily seen from Fig. 5 that inequality $\text{cr}(G) \leq 2(s - 4)$ holds for $s \geq 5$.

To prove $\text{cr}(G) = 2(s - 4)$, $s \geq 5$, it is sufficient to prove that every envelope adds two crossings. This will be shown by means of induction on s . For $s = 5$ assertion holds regarding Lemma 1. By way of induction suppose that $\text{cr}(G) = 2(s - 4)$ holds for every $s < k$. It can be easily seen that by increasing s by one, a new envelope appears in one component of G' . Let G_n denote graph $A_{2,1}^3 \times P_n$. It is clear from Fig. 5 that $\text{cr}(G_k) \leq \text{cr}(G_{k-1}) + 2$, but $\text{cr}(G_k) \geq \text{cr}(G_{k-1})$, since


 Fig. 6. A drawing of graph $A_{7,1}^3 \times P_5$

every two envelopes are edge disjoint. Let G_k^p be a graph created from the graph G_k by replacing all crossings with new vertices in the drawing D , $\text{cr}_D(G_k) = 2(k - 4)$. Let G_k^q be a graph created from the graph G_{k-1}^p by adding vertices $V(G_k) \setminus V(G_{k-1})$ and edges $E(G_k) \setminus E(G_{k-1})$. It can be proved that $\text{cr}(G_k^q) = 2$ in a similar way as proof of Lemma 1. Then $\text{cr}(G_k) = \text{cr}(G_{k-1}) + 2$ immediately follows. ■

ACKNOWLEDGMENT

The author thanks Michal Staš, Daniela Kravecová and especially Marián Klešč for many fruitful discussions, and for reading earlier drafts of the paper.

REFERENCES

- [1] S. N. Bhatt and F. T. Leighton, "A framework for solving vlsi graph layout problems," Cambridge, MA, USA, Tech. Rep., 1983.
- [2] F. T. Leighton, *Complexity issues in VLSI: optimal layouts for the shuffle-exchange graph and other networks*. Cambridge, MA, USA: MIT Press, 1983.
- [3] F. Shahrokhi, O. Sýkora, L. A. Székely, and I. Vrt'o, "Intersection of curves and crossing number of $c_m \times c_n$ on surfaces," *Discrete & Computational Geometry*, vol. 19, pp. 237–247, 1998, 10.1007/PL00009343.
- [4] M. R. Garey and D. S. Johnson, "Crossing number is NP-complete," *SIAM Journal on Algebraic and Discrete Methods*, vol. 4, no. 3, pp. 312–316, 1983.
- [5] P. Hliněný, "Crossing number is hard for cubic graphs," *J. Comb. Theory Ser. B*, vol. 96, pp. 455–471, July 2006.
- [6] R. Fleischer and C. Hirsh, "Drawing graphs," M. Kaufmann and D. Wagner, Eds. London, UK: Springer-Verlag, 2001, ch. Graph drawing and its applications, pp. 1–22.
- [7] I. G. Tollis, G. Di Battista, P. Eades, and R. Tamassia, *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall, Jul. 1998.
- [8] P. K. Jha and S. Devisetty, "Orthogonal drawings and crossing numbers of the kronecker product of two cycles," *Journal of Parallel and Distributed Computing*, vol. 72, no. 2, pp. 195 – 204, 2012.
- [9] K. W. Tang and S. A. Padubidri, "Diagonal and toroidal mesh networks," *IEEE Trans. Comput.*, vol. 43, pp. 815–826, 1994.
- [10] P. K. Jha, "A counterexample to tang and padubidri's claim about the bisection width of a diagonal mesh," *IEEE Transactions on Computers*, vol. 52, pp. 676–677, 2003.
- [11] K. W. Tang and R. Kamoua, "An upper bound for the bisection width of a diagonal mesh," *IEEE Trans. Comput.*, vol. 56, pp. 429–431, March 2007.
- [12] B. A. Pearlmutter, "Doing the twist: Diagonal meshes are isomorphic to twisted toroidal meshes," 1996.
- [13] J. M. Anderson, W. S. Coates, A. L. Davis, R. W. Hon, I. N. Robinson, S. V. Robison, and K. S. Stevens, "The architecture of faim-1," *IEEE Computer*, 1987.
- [14] P. M. Weichsel, "The kronecker product of graphs," *Proceedings of the American Mathematical Society*, vol. 13, no. 1, 1962.
- [15] M. Farzan and D. A. Waller, "Kronecker products and local joins of graphs," *Canad. J. Math.*, vol. 29, pp. 255–269, 1977.
- [16] P. K. Jha, "Kronecker products of paths and cycles: Decomposition, factorization and bi-pancyclicity," *Discrete Mathematics*, vol. 182, no. 1–3, pp. 153 – 167, 1998.

- [17] A. Bottreau and Y. Métivier, “Some remarks on the kronecker product of graphs,” *Information Processing Letters*, vol. 68, no. 2, pp. 55 – 61, 1998.

Overview of Literate and Elucidative Programming

¹Milan NOSÁL (1st year)

Supervisor: ²Jaroslav PORUBÄN

^{1,2}Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹milan.nosal@tuke.sk

Abstract—Software comprehension is often required for maintenance and many times in development process too. For comprehension of the program's code there is support in form of internal documentation. This documentation can be written simply as comments, but literate and elucidative programming brings more sophisticated tools for writing internal documentation. This paper provides overview of these two paradigms.

Keywords—Literate programming, Elucidative programming, Program documentation

I. INTRODUCTION AND MOTIVATION

Very important phase of software development cycle is maintenance and evolution. In these phases there may be many changes in teams configurations and the code written by one programmer has to be understood and updated by another. In the development process, developer communicates with customers to get all the requirements for the software system. These requirements are the basis for the design. But when it comes to implementation of the system, there is an abstraction gap between design and implementation [2]. The design documents express developers ideas and intentions, but the code is written to comply with abstractions of used programming language. Comprehending of such code is a challenge even for senior programmers. Smith [11] states that maintenance programmers spend approximately half of their time simply trying to understand the function of program code.

Changes in teams cause same need for comprehending program code as maintenance. This problem of software comprehension costs lot of resources and therefore is one of greatest needs of software engineering.

This problem is caused by lack of communication of problem understanding. During implementation programmers create their own mental model of problem they solve and apply it in the code. Unfortunately, this mental model is many times lost, because programmers find writing internal documentation (internal documentation explains the source code) too demanding with very low return value. Truth is, writing documentation is not an easy task. One has to be able clearly express his thoughts in required manner, usually in form of text documents or structured diagrams. There is a need to motivate programmers to write documentation during the implementation, while their mental models are clear. Tool or paradigm support can lower demands to write simple and clear documentation. This paper concerns about elucidative programming and its predecessor literate programming, two programming styles that try to tackle this problem.

II. LITERATE PROGRAMMING

"Let us change our traditional attitude to the construction of programs: Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to humans what we want the computer to do."

(Donald E. Knuth, 1984)

Literate programming came as the possible solution to the software comprehension problem described above. Literate programming was invented by Knuth in the early 1980s as a solution to the software comprehension problem [11]. Literate programming tries to change view on the source code. Instead of writing source code to comply with execution platform, literate programming aims at human view of the program and its purpose. The system code is written as a document that follows and explains programmers intentions rather than writing source code and adding comments tries to explain programmer's intentions. The main purpose of using literate programming is to make source code readable, understandable and maintainable. [1]

Literate programming keeps in mind, that source code is usually processed by two different entities. One is source code compiler, that compiles codes into executable program or library. The other one is a programmer that wants to understand the source code written by someone else, or by himself but a long time ago. Therefore the source code is processed in two different paths as it is outlined in fig. 1. Special literate programming language is used to define .web source code (typical literate program is usually referred to as web – consisting of simple parts and relationships put together). This source code is processed by Weave to generate a document for typesetting. Another program, Tangle, is used to produce source code in underlying programming language, for example C. Later it is processed by compiler. [1]

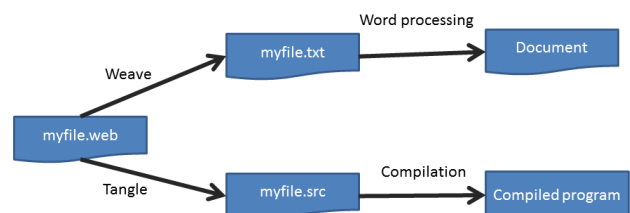


Fig. 1. Processing literate program (source: [1]).

As previous paragraph suggests, literate program is written in at least two languages. Underlying programming language is used to write chunks of code. In the second language pro-

grammar writes description of the code. And there is need for third language to define set of commands to link descriptions and source codes together into one coherent document.

A. Literate programs

Literate program consist of code chunks ([11], Cordes et al. [1] call it modules) of arbitrary granularity. These chunks define relationships of usage between them. One chunk can use multiple other chunks just as a function can use other functions in procedural paradigm.

There are multiple languages that can be used for literate programming. We can mention WEB, literate language developed by literate paradigm author Donald Knuth [1]. This language supports around 30 – 40 commands to define code chunks and distinguish them from documentation. Interesting and probably the most supported now is the noweb, it is very simple to use because it uses only 5 commands. [11]

Following is an example of literate hello world program. One can notice some L^AT_EX commands interleaved by C code. @, <variable>= and <reference> are noweb commands. For more information on noweb language see [9].

```
@
\\section{Hello world}

Today I awoke and decided to write
some code, so I started to write
Hello World in \\textsf C.

<<hello.c>>=
/*
  <<license>>
*/
#include <stdio.h>

int main(int argc, char *argv[]) {
    printf("Hello World!\\n");
    return 0;
}
@
\\section{License}
Later the same day some lawyer reminded
me about licenses.
So, here it is:

<<license>>=
This work is placed in the public domain.
@
```

B. Characteristics of literate programming

The main positive characteristics of literate programming are as follows [11]:

- Maintainability - Code and documentation are written together in the same source document. This integration ensures “active documentation” that evolves and is always consistent with pro-gram code.
- Understandability - Code can be decomposed into smaller chunks and explained in whatever order is most appropriate to aid comprehension. The programmers are not limited to a rigid top-down or bottom-up style; they can use whatever hybrid and combination of formality best fits the exposition. Also background of the problem should be stated.
- Documentation motivation - literate programmers writes their programs as documents. It is easier and more natural to document the program, because literate programming

is documentation-driven (you write documentation with chunks of source code).

- Enhanced readability - Weave program can be used to enhance literate program readability to provide pretty-printing, cross-referencing and so on.

On the other hand there are limitations and negatives of the literate programming, that can hardly be ignored [11], [1]:

- Forcing programmer to write in literate programming language does not guarantee good documentation. One can use literate programming language but code in procedural style. Also quality of documentation cannot be enforced. This concerns attitude of programmers towards writing documentation.
- Literate programming orients on paper-styled documentation. Documentation of literate program is one large document explaining program. Real world programmers do not want to browse whole book to comprehend small part of the program. Weaving is fixed, you cannot try to generate more documents from one web file.
- Literate program is written in single monolithic source file. That is obviously a big limitation to the paradigm. Just try to imagine maintaining such a file for a big real world project.
- Chunks are inflexible, there is always implication that there is a linear series of two-part sections – description and corresponding code chunk. Writing an overview of more chunks is not supported.
- There is a lack of tool support for literate programming.
- And last but not least, lack of guidelines how to author a literate program. So far the literate programming paradigm is accepted and used mainly by scientists. Probable reason may be more complicated code, because literate programmer has to code in three different languages.

III. ELUCIDATIVE PROGRAMMING

Limitations stated in section II-B were evaluated by authors of elucidative programmers as too strong for literate paradigm to be accepted and used in real world projects. Elucidative programming was proposed by the team around Kurt Normark in 2000 [5], [4], [6]. This paradigm is not so far from common programming styles used in real world. Instead of trying to write program in form of one coherent document, classical programming paradigms are used, such as object-oriented programming or structural programming. Program and documentation is written in common manner, in separate documents. To elucidate program source code with documentation, the documentation and source code has to be linked together. In final state, documentation can be read just as in case of literate programming (documentation interleaved with source code chunks), but the proximity between it and source code is not that close – there are only logical links.

As a result, in elucidative programming the source code is left intact without documentation neither surrounding code nor embedded in code. Program understanding is saved in documentation that relates to constituents in source code – physical proximity of literate programming is substituted by navigational proximity. [4]

Source code constituents are source code elements – methods, functions, classes, blocks, and so on. Documentation constituents may be sections or paragraphs.

Normark et al. [4] use HTML and hypertext links for elucidative programming implementation – Elucidator. Documentation is written in form of HTML documents and also source codes are transformed into HTML documents. Navigational proximity is implemented using hypertext links between the documentation and source code. They also support a form of physical proximity allowing programmer to set inline source code that are embedded directly into the documentation [13].

The documentation is read in simple user interface that is outlined in fig. 2. In the documentation pane there are shown documentation pages with links to program. In program pane there are program source code documents. Links from documentation to source code are used to navigate pages in program pane, and links from code to documentation are used to navigate pages in documentation pane. Menu and Index pane serves for providing overview and easy navigation in whole documentation.

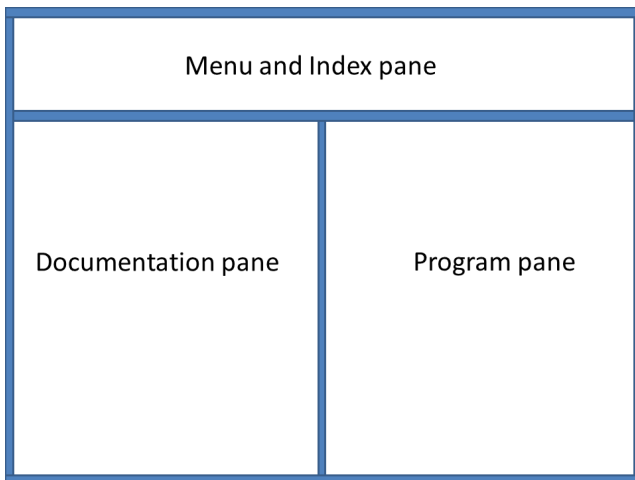


Fig. 2. The layout of panels in Elucidator.

A. Navigation

To navigate from documentation in edoc format (based on HTML) to source code, links use format that is defined by used programming language. For example, following is a link to Java source code:

```
package1.package2.class@method()
```

Similarly, in Scheme a link would look like the following:

```
file\function
```

Sometimes there is a need to navigate to sections that are not named, like blocks in Java. To define such region, comments are used. In Java Elucidator, there are comments with XML-like notation `<e:marker> ...</e:marker>` to define custom navigational section.

B. Java Elucidator

Java Elucidator is elucidative programming implementation on Java platform [6]. Elucidative programming uses two tools, editor and standard internet browser. The editor is used to edit both Java source programs and documentation files in the Edoc language. At first modified Emacs editor was used, but later an integration of tool support in TogetherJ IDE was presented [12]. Browser is used to present the documentation to user.

Java elucidator consists of three main components:

- The abstractor parses and extracts information from the Java classes and the Edoc documentation and passes them to the data model. The data model contains information about all named program and documentation constituents together with their mutual relations.
- The query engine runs queries from the generator on the data model and returns their results.
- The generator is Java servlet on the WWW-server, it dynamically generates current documentation from edoc and java documents and uses the query engine to link them together.

C. Scheme Elucidator

Scheme Elucidator is authored by the same authors as the Java Elucidator. It supports Scheme programming language. On the editor level, the modified Emacs editor is used. In comparison with Java Elucidator, the Scheme Elucidator produces two framed browser as a collection of static HTML pages. [3]

D. Characteristics of elucidative programming

Elucidative programming keeps positive characteristics of literate programming and tackles some of its negative aspects. To be more concrete, following aspects are better in elucidative programming:

- Instead of paper-styled documentation, elucidative programming is oriented to a set of web pages. This format is much easier and natural to browse through.
- Elucidative program is no longer written in monolithic source file. Program is written in multiple source files that conforms to design customs of underlying programming language.
- Due to navigational proximity instead of physical, there is not a problem with chunk's (in elucidative programming they are called source code constituents) inflexibility. Code constituents may be referenced from multiple places in documentation.
- There is no need for any guidelines for authoring elucidative program. Programmer writes code in underlying language in same way, as without elucidative support. All he has to know is to learn how to write links in documentation and marks in the source code.
- Presenting TogetherJ support integration makes elucidative programming more viable to use by real world programmers.
- By supporting program evolution in documentation, elucidative programming makes it much easier to evolve documentation concurrently with software system [14].

But there are still some limitations that haven't been tackled by elucidative programming. The most important one is programmers' attitude to writing documentation. I believe this cannot be tackled by paradigm. This problem has to be dealt with by software engineering educators. Future programmers has to be taught that their programs has to be documented because costs of comprehending software in maintenance are in most cases much larger than for writing documentation.

IV. ATTRIBUTE-ORIENTED PROGRAMMING IN ELUCIDATIVE PROGRAMMING

Attribute-oriented programming (@OP) as a technique for decorating source codes with declarative information boldly influenced the software design [7], programming in general [10] and software configuration [8]. Source code elements can be decorated (marked) with annotations¹. Annotations do not influence the execution of annotated program, but they can be used as configuration technique (more in [8]).

Our original idea was to use annotations as markers of source code for the needs of navigating from documentation to source code. Elucidator uses comments with XML like notation (section III-A). We wanted to substitute these comments with annotations. There are two main reasons that lead us to this decision:

- Writing annotations can be faster and easier than writing comments. Annotations are structured part of Java syntax, and therefore are supported by code completion in IDEs. On the other hand, content of comments is not parsed by IDEs².
- Annotations are directly supported in Compiler Tree API³, API for handling Java abstract syntax trees. Using this API we could easily implement many useful tools or plugins supporting elucidating of sources. One example can be a plugin for IDE to support automatic marking of blocks merely by selecting it with mouse and using keyboard shortcut.

We analyzed the possibilities of annotations usage in this issue. Although it would be very useful enhancement, there is one problem with utilizing this idea. The annotations themselves cannot be used upon unnamed blocks or statements. As this might look like an unbeatable obstacle, there is a workaround to deal with it. Local variable declarations can be annotated. To create annotated blocks, we can add local variable declaration before first and after last statement of the block (or set of statements). These declarations can be annotated to represent starting and ending position of the block to be navigated to. On the other hand, this solution brings up a question whether are annotations needed at all, and whether the needed information cannot be stored in variable declaration itself.

But the need to use local variable declarations makes the whole process of using this approach less effective than we expected. Compiler would optimize the code and would throw away unused declarations. But the code itself would be polluted by the code fragments that do not belong to it. This could have undesired side effects, for example it will make the code hard to read and confusing for anyone who would not know this implementation of elucidative programming, or it may disable code analyzers like PMD⁴ to provide reliable data about code qualities.

To sum up, annotations as they are implemented now did not prove themselves as a useful tool for enhancing navigation to unnamed code blocks in elucidative programming.

¹On Java platform, @OP implementation is called annotations, on .NET there are attributes. We will use the term annotations.

²This is not a rule, for example NetBeans supports code completion for special cases in JavaDoc, but in our case we would need to develop our own plugin to support code completion for our marks. And this solution would be bound to concrete IDE.

³<http://docs.oracle.com/javase/6/docs/jdk/api/javac/tree/index.html>

⁴<http://pmd.sourceforge.net/>

V. CONCLUSION

The paper presents overview of two paradigms for supporting internal documentation. In addition to their descriptions there are stated their main characteristics concerning their applicability. As far as literate programming is originator of the main ideas, elucidative programming makes them more accessible to real world. A brief analysis of the @OP in unnamed code block navigation was presented.

Elucidative programming along with the literate programming helps programmers write internal documentation. Still, one problem stays untackled. The attitude of programmers toward writing documentation. Programmers need to be taught that documentation's value is greater than effort needed to write it. Last semester I have been teaching object-oriented programming and students were required to write interface documentation (documents interface, not the source code) in form of JavaDoc. Almost everyone had complete documentation, but very few of them were of any value. Documenting a method `parseLine()` with sentence "Method `parseLine()`." is absolute nonsense. So not only the programmers has to be taught to write documentation, but to write good documentation. This responsibility is left upon educators in computer science and software engineering.

ACKNOWLEDGEMENT

This work was supported by VEGA Grant No. 1/0305/11 Co-evolution of the Artifacts Written in Domain-specific Languages Driven by Language Evolution.

REFERENCES

- [1] Cordes D., Brown M., "The Literate-Programming Paradigm", in IEEE Computer, Vol. 24, No. 6, 1991, pp. 52–61.
- [2] Greenfield J., Short K., Cook S., Kent S., "Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools", Wiley 2004.
- [3] Normark K., "An elucidative programming environment for Scheme", in Proceedings of NWPER'2000 - Nordic Workshop on Programming Environment Research, May 2000.
- [4] Normark K., "Elucidative Programming", in Nordic Journal of Computing, Vol. 7, No. 2, 2000, pp. 87–105.
- [5] Normark K., "Requirements for an Elucidative Programming Environment", in The Eight International Workshop on Program Comprehension. IEEE Computer Society, June 2000, pp. 119–.
- [6] Normark K., Andersen M. R., Christensen C. N., Kumar V., Staun-Pedersen S., Sorensen K. L., "Elucidative Programming in Java", in The Proceedings on the eighteenth annual international conference on Computer documentation (SIGDOC), 2000.
- [7] Newkirk J., Vorontsov A. A., "How .NET's Custom Attributes Affect Design", in IEEE Software, September 2002, Vol. 19, No. 5, pp. 18–20.
- [8] Porubán J., Nosál M., "Common Abstraction of Configuration from Multiple Sources", Acta Electrotechnica et Informatica, Vol. 11, No. 4, 2011, pp. 25–30.
- [9] Ramsey N., "Noweb – A Simple, Extensible Tool for Literate Programming", <http://www.cs.tufts.edu/~nr/noweb/>, available on 18. 3. 2012.
- [10] R. Rouvoy and P. Merle, "Leveraging Component-Oriented Programming with Attribute-Oriented Programming". Karlsruhe University, 2006, vol. 2006–11, pp. 10–18.
- [11] Smith M., "Towards Modern Literate Programming", Honours Project Report, 2001.
- [12] Vestdam T., "Elucidative Programming in open integrated development environments for Java", in ACM International Conference Proceeding Series 'PPPJ', ACM, Vol. 42, 2003, pp. 49–54.
- [13] Vestdam T., Normark K., "Aspects of Internal Program Documentation - An Elucidative Perspective", in 'IWPC', IEEE Computer Society, 2002, pp. 43–52.
- [14] Vestdam T., Normark K., "Toward Documentation of Program Evolution", in 'ICSM', IEEE Computer Society, 2005, pp. 505–514.

Parallelizing computations over elliptic curve arithmetic in affine and Jacobian projective coordinates

(May 2012)

¹Patrik GALLO (2nd year)
Supervisor: ²Dušan LEVICKÝ

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

²Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

¹patrik.gallo@tuke.sk, ²dusan.levicky@tuke.sk

Abstract— In this paper we present a view over elliptic curve arithmetic with regard to parallel processing on multiprocessor units in general. We provide an analysis of how optimal or not it can be using parallel processors instead of single processor units. Amdahl's law is considered as a first step in this analysis. Analysis includes theoretical introduction to elliptic curve cryptosystem, especially to elliptic curve arithmetic with regard to affine and Jacobian projective coordinate system. Analysis is concluded at the end of the paper.

Keywords— Elliptic curve arithmetic, parallel processing, affine coordinates, Jacobian projective coordinates.

I. INTRODUCTION

The theory of doing computational work in parallel has some fundamental laws that place limits on the benefits one can derive from parallelizing a computation. To understand these laws, we have to first define the objective. In general, the goal in large scale computation is to get as much work done as possible in the shortest possible time. In a perfect world, a computational job that is split up among N processors would complete in $1/N$ time, leading to an N -fold increase in power. However, any given piece of parallelized work to be done will contain parts of the work that *must* be done serially, one task after another, by a single processor. This part does *not* run any faster on a parallel collection of processors (and might even run more slowly). Only the part that can be parallelized runs as much as N -fold faster.

In case of this research paper, the object of parallelism is a part of computation over elliptic curve arithmetic. By elliptic curve arithmetic we mean operations on elliptic curve defined over some finite field. Thus, algebraically it means performing operations over several coordinate systems, for example affine, projective, Jacobian, modified Jacobian, Chudnovsky, and so on. The aim of this paper is to analyze a parallelism in affine coordinate system, but the theoretical basis holds for all the coordinate systems.

II. THE ELLIPTIC CURVE DISCRETE LOGARITHM PROBLEM

Given an elliptic curve E defined over a finite field F_q , and two points $P, Q \in E(F_q)$, find an integer l such that $lP = Q$ in E , provided that such an integer exists. The algebraic objects in the ECDLP (elliptic curves over finite fields) are equipped with only one basic operation: addition of elliptic curve points. On the other hand, objects in DLP are equipped with two basic operations: addition and multiplication of field elements. The additional structure present in the DLP has led to the discovery of the index-calculus methods, which have a subexponential running time. Elliptic curves do not possess this additional structure, and for this reason no one has been able to apply the index-calculus methods to the ECDLP (except in very special and well understood cases). This absence of subexponential-time algorithms for the ECDLP, together with efficient implementation of the elliptic curve arithmetic, is precisely the reason that elliptic curve cryptosystems have proven so attractive for practical use.

III. ELLIPTIC CURVE ENCRYPTION AND DECRYPTION PROCESS

As stated in [4], several approaches to encryption/decryption using elliptic curves are analyzed in the literature. A basic encryption process requires that a plaintext message m is encoded as a point P_m on the elliptic curve. The corresponding ciphertext C_m consists of a pair of points given by: $C_m = \{lG, P_m + lP_B\}$, where l is a random positive integer generated by the source A , G is the base point and $P_B = n_B G$ the point representing the public key of the destination B and n_B is the private key of the destination B . The decryption process is given by:

$$C_m \{P_m + lP_B\} - n_B(lG) = P_m + n_B lG - n_B lG = P_m \quad (1)$$

Most of the processing time consumed by the encryption/decryption processes is consumed by the computations related to lG and lP_B . Therefore, increasing the encryption/decryption rates requires to speed-up the computations related to point-multiplication.

IV. PROJECTIVE COORDINATES

Dealing with O , the point at infinity on an elliptic curve can be troublesome using affine coordinates, the usual (x, y) coordinates that we use to define the Weierstrass normal form of an elliptic curve. One easy way to handle this point is through the use of projective coordinates. Projective coordinates encode a point (x, y) with two coordinates in three coordinates (x, y, z) where (x, y, z) represent any point of the form $(x/z, y/z)$. Such projective coordinates are called *standard projective coordinates*. In particular, we can represent a point on an elliptic curve $P=(x, y)$ as $(x, y, 1)$ and the point at infinity can be represented by the coordinates $(0, 1, 0)$. We can also easily convert from projective coordinates (x, y, z) where $z \neq 0$ into affine coordinates $(x/z, y/z)$.

In addition to being an easy way to handle the point at infinity, projective coordinates are often useful in performing computations on elliptic curves because it is possible to add two points on an elliptic curve using projective coordinates without performing any divisions, which are typically very expensive computationally in finite fields. In cryptographic applications where we may want to perform operations in F_q for large values of q , determining the inverse of an element of finite field can be fairly expensive relative to multiplication and using projective coordinates will often provide a performance advantage over using affine coordinates. *Jacobian projective coordinates* encode an affine point $(x/z^2, y/z^3)$ as the projective point (x, y, z) . In general, each type of projective coordinates requires a different number of field operations to add or double points which is summarized in the table 1. The choice of the most efficient projective coordinate system will depend on the application.

V. ELLIPTIC CURVE ARITHMETIC

Additions are controlled by the following rules:

$$\begin{aligned} O &= -O \\ P(x, y) + O &= P(x, y) \\ P(x, y) + P(x, -y) &= O \end{aligned} \quad (2)$$

A. Point addition in affine coordinates

The addition of two different points on the elliptic curve is computed as shown below:

Input: $P_1(x_1, y_1), P_2(x_2, y_2)$

Output: $P_3(x_3, y_3)$

- | | |
|--|--|
| <ol style="list-style-type: none"> 1. If $x_1 = x_2$ return O 2. If $P_1 = P_2$ then 3. If $y_1 = 0$ return O 4. Else $m \leftarrow \frac{3x_1^2 + a}{2y_1}$ 5. Else $m \leftarrow \frac{y_2 - y_1}{x_2 - x_1}$ 6. $x_3 \leftarrow m^2 - x_1 - x_2$ 7. $y_3 \leftarrow m(x_1 - x_3) - y_1$ 8. Return $P_3(x_3, y_3)$ | $\left. \begin{array}{l} \text{1. If } x_1 = x_2 \text{ return } O \\ \text{2. If } P_1 = P_2 \text{ then} \\ \text{3. If } y_1 = 0 \text{ return } O \end{array} \right\} \text{ Sequential part of algorithm}$
$\left. \begin{array}{l} \text{4. Else } m \leftarrow \frac{3x_1^2 + a}{2y_1} \\ \text{5. Else } m \leftarrow \frac{y_2 - y_1}{x_2 - x_1} \\ \text{6. } x_3 \leftarrow m^2 - x_1 - x_2 \\ \text{7. } y_3 \leftarrow m(x_1 - x_3) - y_1 \\ \text{8. Return } P_3(x_3, y_3) \end{array} \right\} \text{ Part that can be partly parallelized}$ |
|--|--|

where m is the slope of the line intersecting the points on elliptic curve. Because of space shortage I decided not to derive the equations shown above, so you can find it in literature [1].

The addition of a point to itself (point doubling) on the elliptic curve is the case when:

$$P_1(x_1, y_1) + P_1(x_1, y_1) = P_3(x_3, y_3) \quad (3)$$

B. Point addition in Jacobian projective coordinates

The addition of two different points on the elliptic curve is computed as shown below:

Input: $P_1(x_1, y_1, z_1), P_2(x_2, y_2, z_2)$

Output: $P_3(x_3, y_3, z_3)$

- | | |
|---|---|
| <ol style="list-style-type: none"> 1. $u_1 \leftarrow x_1 z_2$ 2. $u_2 \leftarrow x_2 z_1^2$ 3. $s_1 \leftarrow y_1 z_2^3$ 4. $s_2 \leftarrow y_2 z_1^3$ 5. If $u_1 = u_2$ 6. If $s_2 \neq s_1$ 7. Return $(0, 1, 0)$ 8. Else 9. Return (x_1, y_1, z_1) 10. $h \leftarrow u_2 - u_1$ 11. $r \leftarrow s_2 - s_1$ 12. $x_3 \leftarrow r^2 - h^3 - 2u_1 h^2$ 13. $y_3 \leftarrow r(u_1 h^2 - x_3) - s_1 h^3$ 14. $z_3 \leftarrow h z_1 z_2$ 15. Return $P_3(x_3, y_3, z_3)$ | $\left. \begin{array}{l} \text{1. } u_1 \leftarrow x_1 z_2 \\ \text{2. } u_2 \leftarrow x_2 z_1^2 \\ \text{3. } s_1 \leftarrow y_1 z_2^3 \\ \text{4. } s_2 \leftarrow y_2 z_1^3 \end{array} \right\} \text{ Part that can be partly parallelized}$
$\left. \begin{array}{l} \text{5. If } u_1 = u_2 \\ \text{6. If } s_2 \neq s_1 \\ \text{7. Return } (0, 1, 0) \\ \text{8. Else} \\ \text{9. Return } (x_1, y_1, z_1) \end{array} \right\} \text{ Sequential part of algorithm}$
$\left. \begin{array}{l} \text{10. } h \leftarrow u_2 - u_1 \\ \text{11. } r \leftarrow s_2 - s_1 \\ \text{12. } x_3 \leftarrow r^2 - h^3 - 2u_1 h^2 \\ \text{13. } y_3 \leftarrow r(u_1 h^2 - x_3) - s_1 h^3 \\ \text{14. } z_3 \leftarrow h z_1 z_2 \end{array} \right\} \text{ Part that can be partly parallelized}$ |
|---|---|

te the following. High-contrast line figures and tables should be prepared with 600 dpi resolution and saved with no

VI. PARALLEL COMPUTING AND AMDAHL'S LAW

One should expect that most applications will use both CPUs and GPUs, executing the sequential parts on the CPU and numerically intensive parts on the GPUs. In what manner and what range we can parallelize our algorithm gives us answer an Amdahl's law. It considers that in order to improve performance (time to complete the process), only a part of the entire process is usually improved. It shows that the improvement is limited by the fraction of time the improved part is used. It can be stated in terms of two parameters: f , the fraction of time spent in the improved part, and the speedup S , how many times faster the improved part performs. Amdahl's Law gives how many times faster the entire process will be as:

$$S_{\text{overall}} = \frac{1}{(1-f) + \frac{f}{S}} \quad (4)$$

Using absolute execution times, Amdahl's law is in terms

of T_{after} , the execution time after an improvement; $T_{improved}$, the execution time affected by the improvement S , how many times faster the improved part runs, or its *speedup*; and $T_{unaffected}$, the execution time unaffected by the improvement. In these terms, Amdahl's Law states that:

$$T_{after} = \frac{T_{improved}}{S} + T_{unaffected} \quad (5)$$

The thing that is very important to say is that the operations shown in graphs are not equal in terms of computational time and performance. So before we compute the exact speedup and time consumption of parallel computation we have to evaluate the particular operations.

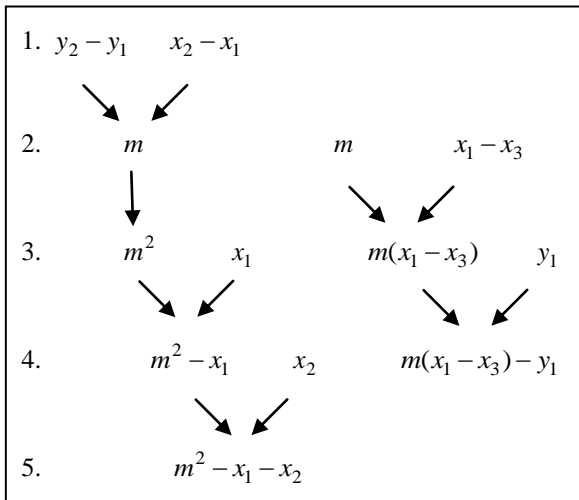


Fig. 1. Point adding dependency graph in affine coordinates

Three operations are mentioned in affine coordinates on the other hand only two operations are listed in projective coordinates. The one extra operation in affine coordinate system is inversion (division). Properties of the inversion are mentioned in literature [2].

A. Parallel implementation – affine coordinates

In point adding graph (Fig 1.) there are exactly three sets of operations that are parallelized, so instead of 9 operations in sequential algorithm we have 5 steps that multicore device needs to run. Let's evaluate the whole process. In first step we make three operations in one cycle instead of doing three separate operations in three different cycles so reached 3 times faster algorithm at this point. However calculation of the slope includes one inversion that cannot be parallelized with other operations because the calculation of point coordinates are dependent with the slope m . In the other two steps (3., 4.) we get twice 2 times faster algorithm in comparison of sequential variant. In terms of performance we reached 7 from 9 operations in parallel. That means using 3 cores at the moment at most, from which we get:

$$S_{overall} = \frac{1}{(1-f) + \frac{f}{S}} = \frac{1}{(1-0.7777) + \frac{0.7777}{3}} = 2.0470 \quad (6)$$

reduction in size should not have an adverse affect the quality of the image.

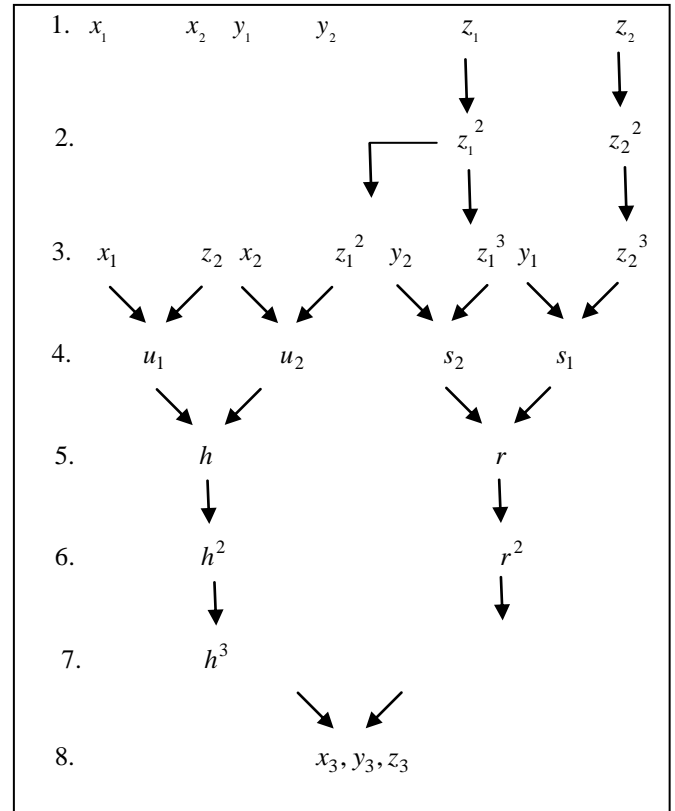


Fig. 2. Point adding dependency graph in projective coordinates

Speedup overall system in adding point algorithm is 2.047 times faster than it would be in sequential execution of program. It is necessary to say that speedup equation doesn't regard the fact that inversion cost 80 times more than simple addition or multiplying. However, this fact is very important in means of parallel computing. Let's see what is the real time needed to perform this algorithm:

$$T_{after} = \frac{T_{improved}}{S} + T_{unaffected} = \frac{7}{3} + 81 = 83.3333 \quad (7)$$

From 88 cycles (80 for inversion 8 for addition and multiplying) we get 83.333 which is just 5.303 % improvement compared to sequential executing program.

B. Parallel implementation – Jacobian projective coordinates

In point adding graph there are exactly two sets of operations that are parallelized, so instead of 14 operations in sequential algorithm we have 7 steps that multicore device needs to run. Let's evaluate the whole process. In first step we make squaring of two coordinates which is done in one cycle. Another cycle is similar and does the same with the same couple of coordinates. In another step parallel algorithm makes four operations in one cycle, which is the biggest speedup in this program. Four other steps are quite simple and are parallelized as it is shown on the graph above. In terms of performance we reached 7 from 14 operations in parallel. That means using 3 cores at the moment at most, from which we

get:

$$S_{overall} = \frac{1}{(1-f) + \frac{f}{S}} = \frac{1}{(1-0.8572) + \frac{0.8572}{4}} = 2.8 \quad (8)$$

Speedup overall system in adding point algorithm is 2.8 times faster than it would be in sequential execution of program. Let's see what is the real time needed to perform this algorithm:

$$T_{after} = \frac{T_{improved}}{S} + T_{unaffected} = \frac{8}{2} + \frac{4}{4} + 2 = 7 \quad (9)$$

Expression on the right side of equation denotes that 8 cycles are computed with 2 cores and 4 cycles with 4 cores. From 14 cycles we get 7 which is 50 % improvement compared to sequential executing program.

Because of shortage of this paper point doubling is not considered, but the speedup improvement is computed exactly the same way.

VII. CONCLUSION

As we can see at the example of the parallelism above, addition algorithm over elliptic curve can be easily parallelized in affine and Jacobian projective coordinates as well. Though, before applying parallel algorithms on existing sequential ones, it is necessary to consider what speedups and time improvements we can get from this rearranging. First of all, affine coordinates are not really good in terms of fast computation on sequential devices. As I mentioned before, this is because of inversion operation applied to evaluate the slope of the line intersecting points. However this fact is limiting in parallel computation too and it cannot be bypassed by any other operation. From these results it could be, in most of the cases, contra-productive to use massively parallel devices to evaluate points on elliptic curves in affine coordinates. On the other hand, the parallelism of addition formula over elliptic curve in Jacobian projective coordinates became really effective in terms of speedups and time improvement. The time improvement of this algorithm is equal to 50% just because of parallelism and the fact that inversion is not included in projective coordinates. From the viewpoint of security implementation, elliptic curve cryptosystems are not just considered as most powerful and secure public key cryptosystem, but can get more and more effective regarding to using massively parallel devices.

ACKNOWLEDGMENT

The work presented in this paper was supported by Ministry of Education of Slovak Republic VEGA Grant No. 1/0065/10 (40%). This work is also the result of the project implementation Development of the Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (60%).

REFERENCES

- [1] MARTIN, L. *Introduction to identity-based encryption* (Book style), Norwood, MA: Artech house, 2008.
- [2] KHALI, H., FARAH, A. *Cost-Effective Implementations of GF(p) Elliptic Curve Cryptography Computations*. (Published work style), International Journal of Computer Science and Network Security. VOL.7 No.8, August 2007.
- [3] KIRK, D., HWU, W. *Programming massively parallel processors* (Book style), Elsevier, Burlington, MA, 2005.
- [4] STALLING, W. *Cryptography and network security* (Book style) Prentice Hall; 5 edition, January 2010.
- [5] SANDERS, J., KANDROT, E. *CUDA by example* (Book style), Adison-Wesley, Boston, MA, July 2010.
- [6] GUTUB, A., IBRAHIM, M. AL-SOMANI, T. *Parallelizing GF(p) elliptic curve cryptography computations for security and speed*. King Fahd University of petroleum and minerals. Dahrhan 31261, Saudi Arabia, 2009.

Remote control of Mitsubishi industrial robot

¹Peter PAPCUN (1st year), ²Matej ČOPÍK (2nd year)
 Supervisor: ³Ján JADLOVSKÝ

^{1, 2, 3}Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

¹peter.papcun@tuke.sk, ²matej.copik@tuke.sk, ³jan.jadlovsky@tuke.sk

Abstract — This article discuss about remote control of industrial robotic arm Mitsubishi MELFA RV-2SDB. We are going to control this robot through smart phone with Android operating system. Our system has ability to control robot's endpoint in Cartesian system axes, or control it's joints. Smart phone is used as a joystick, and also it can be used to make or run recorded sequences.

Keywords — Industrial robot, Robot, Mitsubishi, Control, Bluetooth , Wi-Fi, Android, Smart phone.

I. INTRODUCTION

Under the phrase remote control is meant wireless control through Bluetooth, or Wi-Fi. In this document we write about network connections between nodes and methods of their communication. We describe all applications used in our robot control. Our control architecture consists of five network nodes (personal computer, Wi-Fi router, robot controller, robot, Smart phone) and three applications. Finally, we will evaluate remote control advantages

II. DISTRIBUTION OF SYSTEM

Distributed control system (DCS) is a control system, usually production system, process or any dynamic system in which the system elements are not placed centrally but they are distributed, divided into smaller parts, subsystems that are controlled by one or more control devices. Current industrial information and control systems utilize mainly hierarchic (pyramidal) architectures containing physical and logical distribution elements, integration as a whole, open and scalable. Intelligent features have been applied on a large scale recently whereby direct hierarchic relations are turned into network relations. Emergent trends have also started to appear to a great degree, i.e. merging of originally independent systems, which can result in their new features generation as a whole. [4]

You can see hierarchic architecture of described system on Fig. 1. This architecture does not include all levels of DCS. You can read [4] for more information about DCS. This system pertain a program distribution. Program distribution divides control software to more control units. Software is distributed into robot controller, personal computer and smart phone. Each of them plays specific role in this system.

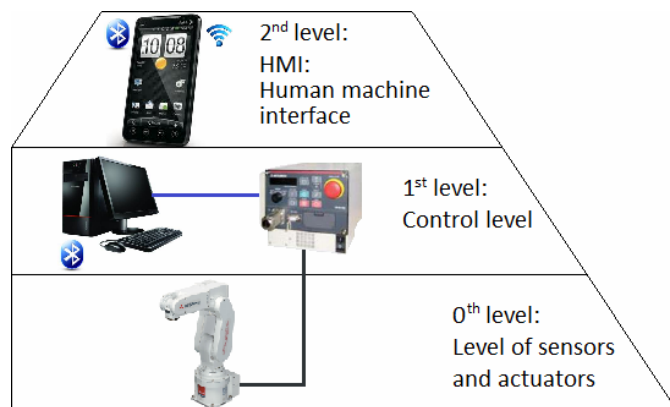


Fig. 1: Hierarchic architecture of described system.

III. HARDWARE AND NETWORK CONECTIVITY

Diploma thesis [1] present control of robot integrated in flexible assembly company. This control system is also divided into three applications. The first application recognizes images. It operates on personal computer (PC). The second program controls whole production of flexible assembly company. Application operates on programmable logic controller (PLC). The third application controls industrial robot and operates on robot controller (RC). In this article we devote robot separately (robot is not integrated into flexible assembly company here).

We will describe used hardware and network connectivity in this part of document. Robot is connected to RC with machine cables. RC is connected to Wi-Fi router through Ethernet cable. PC uses same router as well. You can see a network connection structure of all hardware parts in Fig. 2.

You can find a more accurate description of robot and RC in [1], [2]. PC has Bluetooth device. In PC is installed operating system (OS) Windows 7. Smart phone (SP) have Bluetooth and Wi-Fi devices integrated with OS Android.

Robot is programmed by integrated development environment RT Toolbox 2 with programming language MELFA V (MELFA IV can be also used). RT ToolBox 2 communicates with RC through USB, TCP/IP, or RS – 232 interfaces. We use TCP/IP interface, in this case. Earlier we connected to RC by USB interface, because we had to change default robot IP address to static network address. Then we disconnect USB and we tried to connect TCP/IP (Ethernet cable). Connection was successful. PC already belongs to network. We configured Wi-Fi router, so that other devices can connect to it using Wi-Fi.

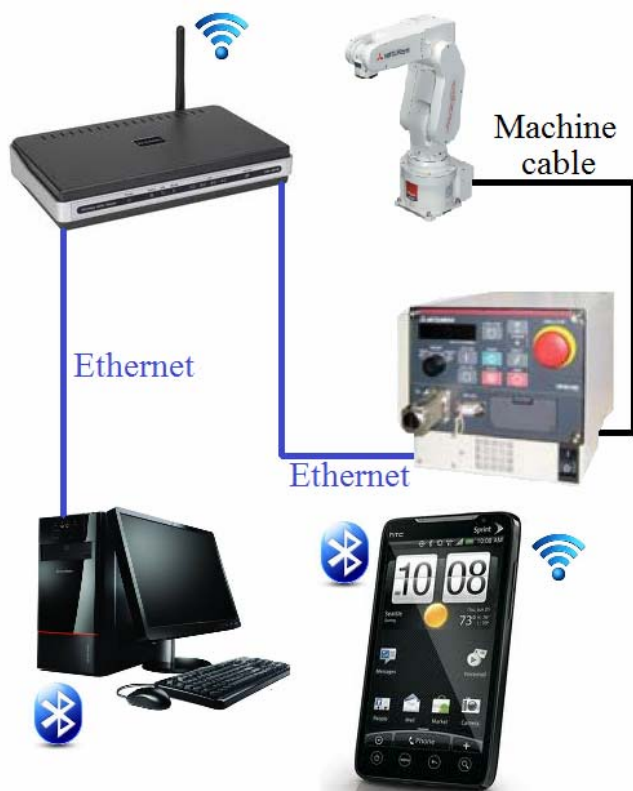


Fig. 2: Network connectivity.

IV. SOFTWARE

Robot control consists of 3 separate applications:

- Program in RC,
- Program in PC,
- Program in SP.

Program in RC

The simplest algorithm is applied in RC:

```
Open "ENET:192.168.0.2" As #1
Mxt 1,1,50
Hlt
End.
```

Command MXT is one of loop commands. This command is specific. It is expecting User Datagram Protocol (UDP) packet from specific IP address. You can see three MXT arguments. The first argument is communication channel number. The second argument is position data type (Cartesian system, joint, motor pulse). The last argument is filter time constant. Flowchart for previous algorithm looks is on Fig. 3.

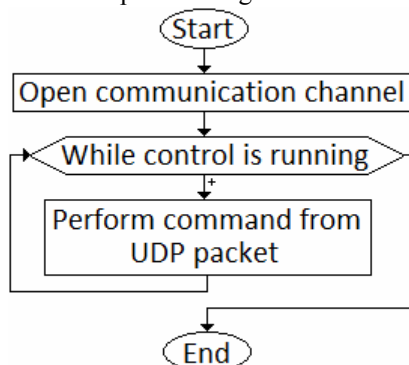


Fig. 3: RC flowchart.

Data of UDP packet (without Ethernet, IP, UDP headers and Ethernet Trailer) consists for example of: command,

transmission and reply data type designation, position data, timeout time counter value, etc. Structure of UDP packet (its data part) is determined by Mitsubishi. When RC receive UDP packet, robot will move accordingly to packet instruction. In other words, an MXT is command for real time control of robot.

PC application

We used C# programming language. The following figure (Fig. 4.) shows the high level app diagram:

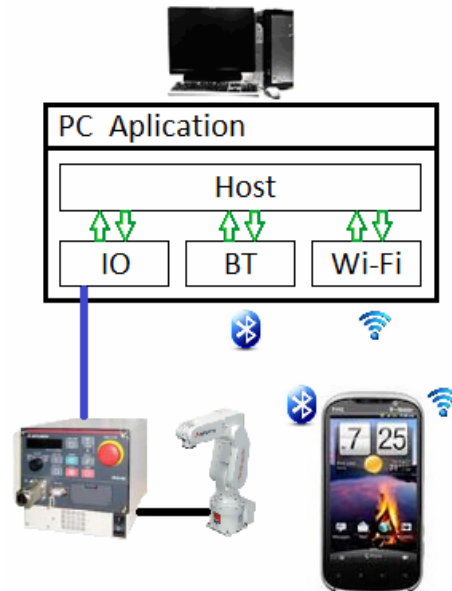


Fig. 4: Program diagram.

Main and basic application module is Host. You can load plug-in modules to host. IO, BT and WiFi are plug-in modules. Plug-ins implements custom plug-in interface in order to be “plugged into” host module.

The main task of Host is to maintain plug-ins in memory and inform them about events, which user invokes in main application (language change, request plug-in information, etc.). Another task of Host is switching between displayed plug-ins, as required by user. When user has switched between plug-ins, instances of plug-ins run in background and their functionality is not compromised in any way.

IO is next plug-in module. Its task is communication with robot. The IO module obtains information about robot (arm position, joint rotation, etc.) and this information is passed over through host to special global class. This class maintains current information about robot. Main task is to execute orders from command queue (type FIFO – first in first out) which is also part of global class. Other plug-ins write command to command queue and IO takes care of sending and performing those commands in robot.

Plug-ins BT and WiFi communicate between SP and plug-in IO. BT communicates with Bluetooth interface. We are using standard communication protocol RFCOMM. WiFi communicates through TCP/IP interface. We are using UDP in WiFi. This plug-ins behaves as server, which are waiting for client to connect through Bluetooth interface or WiFi. Application obtains connection information about remote client, after accepting connection. Then communication is running between SP and PC. You can control robot directly. Robot coordinates are sent to SP (by IO plug-in) and SP is

sending new coordinates. You can record and activate command sequence from SP also. Mobile phone sends command for sequence to module BT or WiFi. Or you can teach new sequences. By mobile you confirm some points, these points are saved to file in computer and then you can activate those new sequences from file.

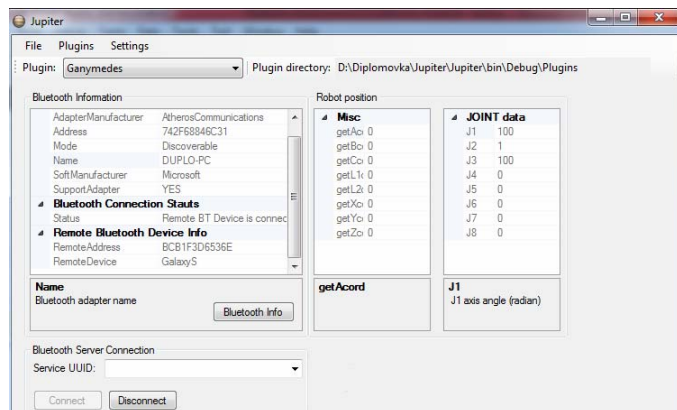


Fig. 5: PC application.

Application in SP

App is programmed in Java programming language using integrated development environment (IDE) Eclipse Galileo 3.5.2. Our app is then sent to SP through this IDE. You can see basic and the first screen of application on figure (Fig. 6).



Fig. 6: Basic screen of program in SP.

RobotControl Activity (Fig. 7) is main activity. User can receive information from PC about robot position through this activity. User can define every coordinate independently. He can choose position data type (Cartesian system, joint, motor pulse), which he sends to PC. PC (module IO) sends this information to RC. Manual control is next option of by arrow buttons (Fig. 7).

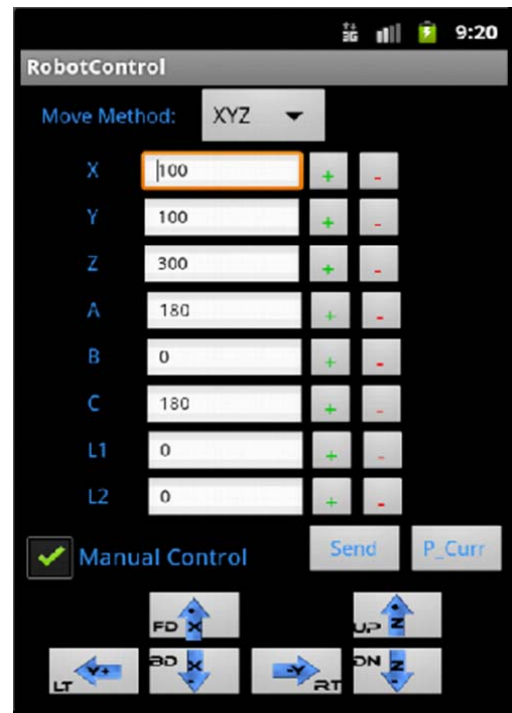


Fig. 7: RobotControl Activit (SP screen).

User activates sequences from next part of app. List of sequences is on screen called with same name. Those sequences are saved in PC (in files). User has option to teach robot new sequences, with other part of application. User confirms some points (in Cartesian system), this points are saved to file in computer. Then users can active new sequence through Sequences activity.

V. CONCLUSION

We tried this type of control. Every part of applications are working. SP replace Teaching pendant, but not completely. Because user can program robot with Teaching pendant.

With SP we can:

- control robot with arrow buttons,
- send robot to a chosen coordinate,
- rotate joints to a chosen axis,
- active sequences,
- teach robot new sequences.

With SP we can not:

- program robot,
- teach robot a sequences with define type of movement.
- set RC parameters.

We can program application, which can do mentioned things (without set RC parameters). Purpose was not to create remote Teaching pendant in SP. We are only trying option of controlling robot in C# with real time control (command MXT in RC). For example one of diploma thesis in this year (in our department) is about real time control of robot through cameras system with infra sensor. You can control robot with your body movement in application of this diploma thesis.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

This work has been supported by the Scientific Grant Agency of Slovak Republic under project Vega No.1/0286/11 Dynamic Hybrid Architectures of the Multiagent Network Control Systems.

REFERENCES

- [1] PAPCUN, P. 2011. Control of robot integrated in flexible production line, diploma thesis, Košice, Slovakia, 2011.
- [2] Instruction manual, CRnQ/CRnD Controller, Mitsunishi Electric, Ratingen, Germany, 2010.
- [3] Product leaflets, RV-2SDB, Mitsubishi Electric, Ratingen, Germany, 2010.
- [4] JADLOVSKÝ, J. – LACIŇÁK S. – CHOVAŇÁK J. - ILKOVIČ J. 2010. Proposal for distributed control system of flexible production line, Journal of Cybernetics and Informatics, vol. 11, Košice, Slovakia, ISSN: 1336-4774
- [5] ILKOVIČ, J. – ČOPÍK, M. – JADLOVSKÝ, J. – LACIŇÁK, S. 2011. Technological level of flexible manufacturing system control, Acta Electrotechnica et Informatica, vol. 11, no. 1, pages 20 – 24, Košice, Slovakia, ISSN: 1338-3957
- [6] ILKOVIČ, J. – ČOPÍK, M. 2011. The assembly line model at Department of Cybernetics and Artificial Intelligence, SCYR 2011: 11th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice: proc. - Košice : FEI TU, 2011, pages 373-376, Košice, Slovakia, ISBN 978-80-553-0644-5

SCADA/HMI System in Action Graphs

¹Martina ĽALOVÁ (3st year), ²Lukáš LACINÁK (1st year)

Supervisor: ³Valerie NOVITZKÁ

¹Dept. of Computers and Informatics, FEEI TU of Košice, Slovak Republic

²Dept. of Cybernetics and Artificial Intelligence, FEEI TU of Košice, Slovak Republic

³Dept. of Computers and Informatics, FEEI TU of Košice, Slovak Republic

¹martina.lalova@tuke.sk, ²lukas.lacinak@tuke.sk, ³valerie.novitzka@tuke.sk

Abstract—In this contribution we want to present the architecture SCADA/HMI (Supervisory Control and Data Acquisition / Human Machine Interface) in Action Graphs focusing on communication of processes. Architecture of SCADA/HMI is considered to be final, deterministic and closed system.

Keywords—SCADA/HMI, action calculus, action graph

I. INTRODUCTION

SCADA/HMI systems are frequently used in industry for Supervisory Control and Data Acquisition of industrial processes and as Human Machine Interface [1]. This is an often used and complex architecture with a complex internal structure of processes. The complexity of relations between processes is difficult to describe in computer science. We use the action calculus as instrument to describe syntax and action graphs for semantics [2], [3].

II. SYSTEM AND ACTION CALCULUS

System is possible to describe as a formal transcript of behavior our architecture [4],[5], [6]. This transcript defines interaction between processes and designates specification [7]. Knowing the properties of system and interactions between its components leads to the knowledge of the system. Knowing the composition of the system allows us to understand the structure and its architecture. Let consider understanding of the structure as the understanding of its components and their interactions as partly closed autonomous system.

Communication takes connection between at least two processes. It is needed for establishing a connection, that one process initiates the connection. The connection is initiated by sending a message that another or other processes may receive. For example, process *A* sends message through port *b* to the process *B* (connected ports have always the same names).

Processes send and receive messages through their ports, which are synchronized. The messages contain data.

The action calculus can be defined as a formal method for the description communicating processes and analysis of its properties.

The process is a series of actions, status changes or functions which lead to results.

Action is the smallest unit of system's activity. Complex behavior of a system by means of action graphs can be expressed by composition of actions. The action is considered to be the process..... happening in the system, which will represent the smallest unit of representation.

A. Action calculus

The action calculus provides a framework for describing models of interactive behavior, where the graph corresponds to the process and dynamics of the graph corresponds to the interaction between them [8], [9].

Action calculus:

$$P \stackrel{\text{def}}{=} x < a > | \text{def } D \text{ in } P | P \circ P | \mathbf{0} | go < \alpha, \kappa > | \text{halt}() | \text{fail} < \alpha, \kappa > | D \stackrel{\text{def}}{=} J \triangleright P | \alpha[D : P] | \Omega\alpha[D : P] | \mathbf{T} | J \stackrel{\text{def}}{=} x \langle a \rangle | J | J$$

behavior of syntax units

- $P \stackrel{\text{def}}{=} x < a >$ - process *P* with port *x* and message *a*
- $\text{def } D \text{ in } K$ - subelement *D* in element *K*
- $P \circ P$ - define parallel composition of processes
- $\mathbf{0}$ - define null process
- $go < \alpha, \kappa >$ - define migration of connection
- $\text{halt}()$ - action of stop
- $\text{fail} < \alpha, \kappa >$ - detection of errors in connection
- $R \stackrel{\text{def}}{=} J \triangleright P$ - local rule to designate *R*, reaction *J* of action of process *P*
- $\alpha[D : P]$ - define sublocation α , where *P* is subelement of *D*
- $\Omega\alpha[D : P]$ - dead sublocation, inactive or exhausted sublocation
- \mathbf{T} - inactive element, terminal definition
- $J \stackrel{\text{def}}{=} x \langle a \rangle$ - message pattern on connection
- $J | J$ - join-pattern

Separation of components of the system *S* in a multi-layer client-server SCADA/HMI architecture and management of the system into layers increases its performance, as there is a possibility of parallel processing supported by the servers, as well as good communication speed between the components. From the point of functional and logical arrangement of processes it is needed for action graphs to create two separate components *K*₁ and *K*₂.

We define a complete system *S* as

$$\text{def } K \text{ in } S \quad (1)$$

and set of components $K \in \{K_1, K_2\}$

B. Operating/communicating component

All controlled parts by the SCADA/HMI system are established for the purpose of management of technological processes. The technological process is a managed and controlled

source of data, which are distributed through communication system for all processes. The actual data as a continuous flow are recorded in the *Database Server*, as you can see on Fig.1. Operating/communicating component creates a complex unit, which forms the core architecture of SCADA/HMI. It allows other processes to access output data of technological process. Technological process is considered as all the processes executed in industrial production. The practical realization of the technological process requires some degree of control. The management processes are covered by the component K_1 . The management activities of the component K_1 are driven by processes of *ApplicationSever*.

These processes perform control actions, which are distributed through ports. General actions are specified by the functions which describe the behavior. Examples of these actions can be actions to stop the implementation *ApplicationSever.halt()* or an error message *fail(int)*, where *int* parameter describes the type of errors [10], [11].

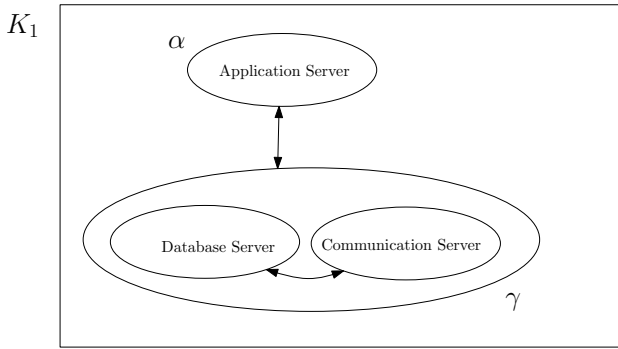


Fig. 1. Operating/communicating component of architecture SCADA/HMI

We can define the sublocation in the system and also within the component by rule

$$\alpha[D : P], \quad (2)$$

where sublocation α with process P in D is defined. This way we can differentiate various sublocation. For example by rule $\Omega\alpha[D : P]$ we can differentiate various dead sublocations, inactive or exhausted sublocations.

In this component K_1 we can define sublocation α . We use rule (2) and we can write

$$\alpha[K_1 : \text{ApplicationServer}]$$

Similarly, we define a dead sublocation

$$\Omega[\gamma : \text{DatabaseServer} \wedge \text{CommunicationServer}]$$

C. Client's component

The process $P \stackrel{\text{def}}{=} x \langle a \rangle . \text{WebClient}$ is from the perspective of system a oneway connection as a terminating process on the *WebClient*. This process P has port x and message a , which defines action of process P .

For this component internal reaction rules can be defined as in Fig.2. The rules defined by prescription

$$R \stackrel{\text{def}}{=} K_2 \triangleright P$$

can be defined by different levels of access rights. These access rights are usually already adjusted in the user accounts. User accounts allow to set the connection type, add connection, set parameters and limit use of the connection [12].

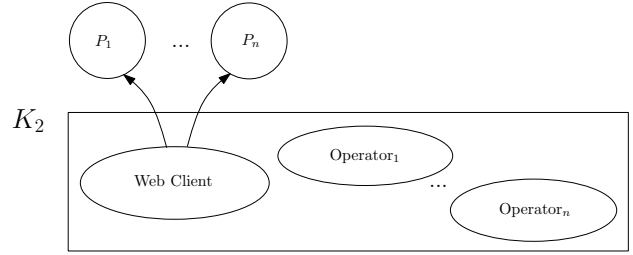


Fig. 2. Client's component of architecture SCADA/HMI

A very important element in the architecture of the system and the action graphs is a connection. The connection is always considered from the perspective of either the system or from the perspective of its users. In this article we consider the connection from the perspective of the system [13], [14].

We consider three basic levels of system architecture. In this architecture the lowest level is the technological process which contain specific process *TouchPanel*.

At this level *CommunicationServer* is used to manage the connection inside of component K_1 . Component K_1 covers *ApplicationServer* and sublocation *DatabaseServer* and *CommunicationServer*.

Into this component are connected two types of developers and next type operators. Operators are regarded as the two types of processes. One type of processes has access to the whole component of K_1 . Second type of processes are not connected directly to the component, but by a process of *ApplicationServer*.

The highest level is the component *client's component* K_2 . This component associates processes (clients), which are connected to other components and have no possibility to influence the activity of other processes, they can only see their outputs, or have limited rights.

The behavior of this architecture we can describe by the composition of actions of individual processes. When we consider that the communication from the process P as user towards the database, then we can follow the composition of actions of individual processes, which are managed by the process *CommunicationServer*.

We write the relationship between the user process P and the *TechnologicalProcess*:

$$P \rightarrow \text{TechnologicalProcess}.$$

However, the user is not able to access directly the technological process.

The user can see data in the database.

We will use the tracking actions a and the necessary processes which were identified by shortcuts:

$$\text{CommS} \langle a.P \circ a.\text{WebC} \circ a.\text{WebS} \circ a.\text{ApplS} \circ a.\text{DbS} \rangle$$

We assume that the data in the database come from the technological process:

$$\text{TechnologicalProcess} \rightarrow \text{DatabaseServer}.$$

So we are able to put our analyzed architecture into a complete action graph, which will represent the whole architecture of SCADA/HMI. Action graph of architecture SCADA/HMI with individual connections is represented on Fig.3

A specific process *TouchPanel* was incorporated into the architecture. This process is considered as a component of

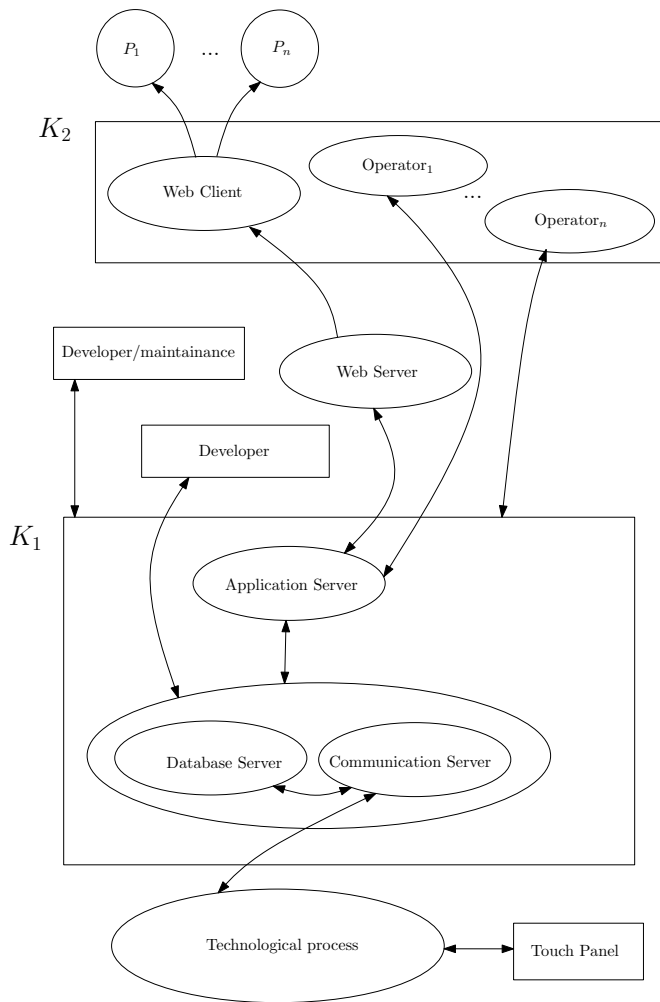


Fig. 3. Architecture SCADA/HMI as complex component system

the simple reason that the panel is directly connected to technological process, but outside of *CommunicationServer*, i.e. outside of the overall management of system. It allows direct communication.

III. CONCLUSION

Creating architecture of controlled system is based on the required complex performance of the system, which are realized by existing technical hardware.

Each hardware has its own specifications which are creating restrictions in the development of architecture. Therefore the system has to be adjusted during the implementation to fulfill the proposed requirements from the available implementation components.

Architecture formed in this way can be slightly deviated from the originally specified requirements. Such variations are hardly recognizable in the system at the implementation level of management.

Action graphs solve the problem from the perspective of abstraction. They provide a model system as a complex set of verification features which originate in the design of control system.

It allows a closer look to individual processes and to the connection between the processes. It allows also to observe the behavior of this process and thus to verify the behavior of the implemented architecture.

Action graphs' approach allows multi-level abstraction. This architecture enables monitoring of the behavior on different levels.

REFERENCES

- [1] S. W. Daneels, A., "What is scada?" In, International Conference on Accelerator and Large Experimental Physics Control Systems, Trieste, Italy, 1999. [Online]. Available: <http://accelconf.web.cern.ch/accelconf/ica99/papers/mcli01.pdf>
- [2] M. H. Andrew Barber, Philippa Gardner and G. Plotkin, "From action calculi to linear logic." Computing Laboratory, University of Cambridge, Cambridge CB2 3QG, Englan, 2002.
- [3] R. Milner, "The space and motion of communicating agents." Lecture Notes for ESSLLI, Department of Mathematics and Statistics, McGill University, 2009.
- [4] M. Ľalová and V. Slodičák, "Some useful structures for categorical approach for program behavior," in *Proceedings of CECIIS 2010 Central European Conference on Information and Intelligent Systems*, Varaždin, Hrvatska, 2010.
- [5] G. L. Cattani, J. J. Leifer, and R. Milner, "Contexts and embeddings for closed shallow action graphs," Tech. Rep., 2000. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.36.7622&rep=rep1&type=ps>
- [6] V. Slodičák, "Some useful structures for categorical approach for program behavior," in *Journal of Information and Organizational Sciences*, Vol. 35, No. 1, 2011, pp. 1846–9418. [Online]. Available: www.jos.foi.hr/http://www.jos.foi.hr/
- [7] M. Franečková and comp., "Safety communication of industrial networks." In Slovak EDIS. ŽU Žilina, 2007.
- [8] R. Milner, "A calculus of communicating systems," Lecture Notes Comp. Science, Springer, Berlin, 1989.
- [9] T. F. Gritzner, "The action graph model as a link between abstract relation algebras and process-algebraic specifications," 2007. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.41.3855&rep=rep1&type=pdf>
- [10] I. Zolotová and L. Laciňák, "Data acquisition and supervisory control communication layer and life cycle of application," Department of Cybernetics and Artificial Intelligence, Technical University of Košice, Letn 9, 042 00 Košice, Slovak Republic.
- [11] C. S. Lazar, C., "A remote-control engineering laboratory," in *IEEE Trans. Ind. Electron.* vol. 55, no. 6, 2009, p. 23682375.
- [12] M. Ľalová, V. Novitzká, and V. Slodičák, "Algebras on the duality principle for program behavior," in *Journal of Information, Control and Management Systems*, 9, 1, 2010, 2010, pp. 61–70.
- [13] A. Malnič, "Action graphs and coverings," University of Ljubljana, Institute of Mathematics, Physics and Mechanics, Department of Mathematics, Jadranska 19, 1000 Ljubljana, Slovenia, 2000.
- [14] M. Ľalová, "Dynamic systems and their description: Calculus vs. action graphs," in *SCYR 2010 - 10th Scientific Conference of Young Researchers FEI TU of Košice*, 2010, p. 270–274.

Solution for receiving live broadcast of lecture with interaction between teacher and students

¹Dávid CYMBALÁK (1st year)

Supervisor: ²František JAKAB

^{1,2}Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹david.cymbalak@cnl.sk, ²frantisek.jakab@cnl.sk

Abstract— Paper deals with a design of solution, which can be used for receiving live streamed lectures with interaction on student and also on teacher side. Based on related analysis, selected existing streaming technologies in cooperation with newly created web components are joined into multimedia solution with educational abilities. Main functionalities of proposed solution are receiving video stream, online voting system and online question system. Video stream is delivered via media server in form of flash file format to flash player located on web page. Online voting system can be used to make survey or small test on viewers of broadcasted lecture. Results of voting are available in real time on teacher side in form of dynamically changing block on teacher's web page. Online question system is designed for ability for viewers to interact in form of text question in real time during the broadcast. Questions from students are delivered to queue located in web page, where can be read by teacher and can be answered directly to the video stream. This functionality brings possibility to adapt further course of lecture based on real time feedback from students and react in real time. Experimental realization of proposed solution was implemented in form of web page, which has been tested on sample group of students.

Keywords—Voting system, Live streaming, E-learning, Web technologies.

I. INTRODUCTION

Interest for comprehensive solutions for e-learning is always growing, because topic of interactive multimedia in education sector is quite extensive. E-learning in conjunction with interactive multimedia is also becoming essential tools of self-education. In general, computer technology greatly assists in the learning process for learners with disabilities, or for learners with special educational needs. Allows better visibility or customization of text documents, playing text as sound, but also facilitate access to learning materials for people who cannot attend lecture [1]. In present it is popular to use streaming technologies such as IPTV for remote access to videos containing learning materials. Usage of IPTV is polyfunctional, but it has the largest representation in the replacement of analogue television broadcasting. Other usage includes the possibility of distant education, which is unlike video conferencing, technologically simpler option of providing one video stream to large number of recipients, but

at the expense of losing interactivity from side of recipients. This fact creates a new dimension in looking at education, where barriers of distance are broken down and there is opportunity to participate in full training process through the network connection anywhere in the world. IPTV adapted to educational purposes appears to be a means, which would be in the future as one of the primary sources of education [2,3]. Despite the progress that these technologies have brought, the student is usually situated in the role of passive receiver of teaching materials without the possibility to active intervention. It is therefore necessary to bring more interaction to multimedia elements, display elements and streaming technologies, what transform passive students into the active members of learning process with possibility of remote interaction.

II. ANALYZE OF PROPER TECHNOLOGIES

A. Streaming Technologies

Streaming is a technology, which ensures the transfer of multimedia content over a network. It allows transmission of audio, video or even other multimedia content from sender to recipients. This technology can be used to real-time broadcasting, but also for receiving video on demand (VOD) [2]. Most common technologies, which have been used in web based streaming solutions is Wowza, Flash Media Live Encoder and JW player.

Wowza Media Server

It is fully interactive server for streaming multimedia content with full support for H.264 and RTMP/ RTMPS [3]. In this solution it will be utilized as the streaming server for live broadcast or also for video on demand service in archive section.

Flash Media Live Encoder

This software is built to stream live video in real time and simultaneously convert to Flash. It supports codec H.264 and On2 VP6. Program can simultaneously live broadcast and save stream to file in FLV format [3]. In this solution it will be utilized to provide encoding live broadcasts from camera.

JW Flash Player

JWplayer is web based flash player with support of RTMP protocol and HTML5. It brings feature called “dynamic bit rate switching”, which set the bit rate of video dynamically according the information from user about bandwidth. JWplayer will be utilized to play live broadcast and video on demand in web interface on client side.

B. Voting Systems

Common voting systems consist of several small hardware voting units and on central unit. Hardware voting units are usually portable and powered by batteries [4]. Portable voting machines have great functionality, but they are quite expensive to use them for large group of students. For non-anonymous voting it has to pair each device to concrete student to ensure voting relevance. For this reason it is less expensive to integrate voting system to current solution in way, where students can vote via web interface after authorization and authentication.

III. DESIGN OF SOLUTION

A. Expected functionalities

Expected functionality of proposed solution can be divided into the following main points:

- Deliver live stream of lecturer speech and image from the camera to students via network.
- Provide archive of historical lectures and possibility to comment lecture or ask question to lecturer offline.
- Allow to ask questions from students to lecturer directly during live broadcast via web interface.
- Allow students voting and provide dynamically results of voting from student to lecturer via web interface.
- Possibility to change streaming video source addresses via web interface.

B. Interconnection between components

Proposed design can be shown as diagram of interconnections between chosen components (Fig 1.). Lecturer is standing behind laptop and is captured by camera. Output from video camera is connected to capture card with HDMI or firewire input. Speech is captured by microphone, which is connected to audio mix. Video and audio inputs are encoding via instance of Flash Media Live Encoder (FMLE). Encoded live stream are delivered as unicast to application of Wowza media server, which ensure optimized multicast streaming reachable from web based flash player – JW. This window of JWplayer is located into code of web page section for Viewers. Viewers section of web portal contains also questions board and voting board. Newly asked questions are storing into database, from where they are read in real time and sending to queue in Lecturer section of portal. Voting is configured in Lecturer section, results of voting is collected to database from each voting board in Viewer section. Voting results are read from database and they are dynamically shown in voting board window in Lecturer section. Archiving of actual lecture is ensured in form streaming to files during

live. Administrator section provides possibility to set addresses of streaming sources or possibility to manage archive section.

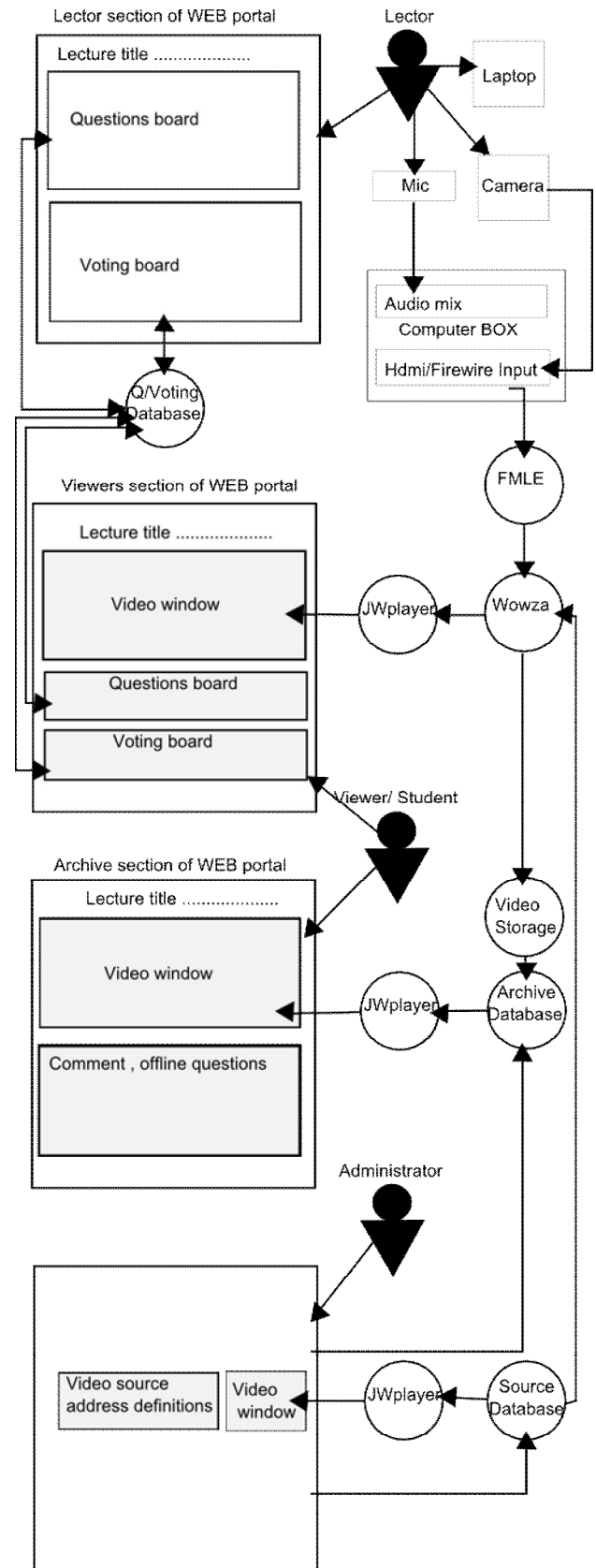


Fig. 1. Design of interconnection between component of proposed solution.

IV. EXPERIMENTAL REALIZATION

Solution design was transformed to experimental realization of web page using the PHP, HTML, JavaScript in cooperation with jQuery UI library and jQuery theme in CSS.

Viewers can watch broadcast in flash player window without interruption during using other services. Also they can write and send questions or vote for alternatives in case of those features are enabled (Fig. 2). State of features, new voting or new alternatives are changing dynamically without need to refresh site. Dynamic results of voting and dynamic queue of questions are located in Lecturer section. The oldest question is on the top of queue and it can be removed by click. New voting can be defined in collapsible block, which will appear after click on related button. Enable or disable voting and question features can be managed by clicking on actual state (Fig. 3). Streaming address and name can be defined in Admin section of portal. Two small windows are showing preview of video source (Fig. 4).

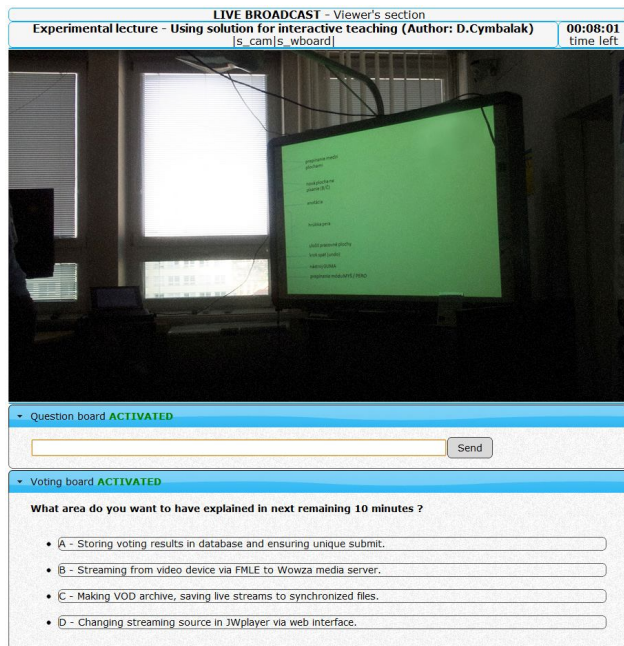


Fig. 2. Viewer section of web page for watching asking and voting

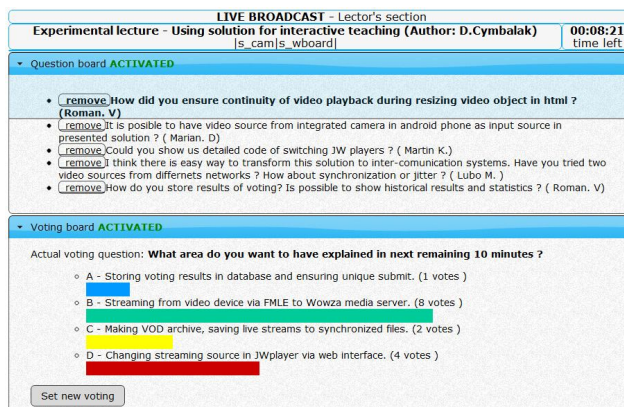


Fig. 3. Lecturer section of web page for set voting and seeing dynamic results



Fig. 4. Admin section of web page for setting streaming options.

Overall solution was tested in form of experimental presentation of video team activities on small group of student members located outside of classroom connected via web page. In comparison with the traditional form of lecture, several improvements were observed in the teaching process. Time was utilized more efficiently due to adaptive reactions to voting or new questions. Based on results teacher can focus on desired topics or eliminate adverse topics. Remote participants had opportunity to immediately respond in form of questions, which are answered by teacher directly to the video. Count of questions had increased, when the questions had been set as anonymous. It could be said, that teacher receives a tool to quickly test the gained knowledge of local or remote participants. Teacher had also possibility to choose order of answering the questions in the queue. In general lecture has become more interesting for participants.

V. CONCLUSION

The concept of described solution brings new perspective of modern education. Overall solution has several ways of interactivity between students and teacher. One possibility of interaction is web section for addressing questions to lecturer, who answers real time in live broadcast. Voting systems ensures integration of unlimited number of students and voting results are displayed dynamically directly to lecturer. All of these functionalities are very helpful for learning process not only for distance learning but also for organizing science conferences or corporate trainings.

ACKNOWLEDGMENT

This work is the result of the project implementation Development of the Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] J.Zounek – K. Sedova, *Teachers and technologies : Between traditional and modern concept*, Brno : Paido, 2009.
- [2] G. O'Driscoll, *Next generation IPTV services and Technologies*, Hoboken: Wiley, 2008.
- [3] D. Cymbalak, *Next generation IPTV*, Diploma thesis, Kosice: Technical University of Kosice, 2011.
- [4] D. Betak - M. Ozvoldova, *Clicker response system – Didactic tool for knowledge verification*, Trnava : Typi Universitatis Tyrnaviensis, 2011

State Control Design for Systems with Time Delay

Daniel GONTKOVIČ (3rd year)
Supervisor: Dušan KROKAVEC

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

daniel.gontkovic@tuke.sk

Abstract—This paper is concerned with the problem of stabilization and state control design for linear dynamic systems with time delay. The problem addressed is presented as an approach based on Lyapunov-Krasovskii functional and on Linear Matrix Inequalities (LMI). The solution is proposed by the bilinear matrix inequality (BMI). A solution for inequality can be found using the iterative procedure if the matrices of Lyapunov-Krasovskii functional are the LMI variables. Simulation results show the feasibility of the proposed approach.

Keywords—Lyapunov-Krasovskii inequality, Bilinear and Linear matrix inequality, Time-delay systems, State control design, Schur complement

I. INTRODUCTION

Systems with delays appear in many industrial applications where time delays may be a source of instability. Therefore new procedures are still being developed to compensate or eliminate time delays.

Over the past few decades a lot of attention has been paid to the problem of stability analysis and controller design for systems with time-delay. Scientists have been studying and developing the use of Lyapunov method for the stability analysis of the time delay systems, especially the pioneer Krasovskii [3]. Nowadays, normally for the stability issue of the time delay systems modified Lyapunov-Krasovskii functionals Shaked et al. [7] can be used to achieve the delay-independent stabilization and these are applied to the controller synthesis and observer design. A overview of the progress in this research area can be also found in Niculescu et al. [6], and the references therein.

The bilinear matrix inequality can be solved with the iterative procedure if the matrices of Lyapunov-Krasovskii functional are the LMI variables. Then the design task of the stabilizing controller for the closed-loop system is transmuted into LMI framework and solved. The paper describes 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 $\beta = \langle 0, 1 \rangle$.

II. SYSTEM MODEL

Equations of linear dynamic systems with time delay are as follows

$$\dot{\mathbf{q}}(t) = \mathbf{A}\mathbf{q}(t) + \mathbf{A}_d\mathbf{q}(t - \tau(t)) + \mathbf{B}\mathbf{u}(t) + \mathbf{B}_w\mathbf{w}(t) \quad (1)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{q}(t) \quad (2)$$

$0 \leq \tau(t) \leq \tau_m < \infty, 0 < \dot{\tau}(t) \leq \beta < 1, \tau_m, \beta \in \mathbb{R}$
 with the initial condition

$$\mathbf{q}(\vartheta) = \varphi(\vartheta), \quad \forall \vartheta \in \langle -\tau_m, 0 \rangle, \quad (3)$$

where $\tau(t) > 0$ is the delay of the state, $\mathbf{q}(t) \in \mathbb{R}^n$ is vector of the state, $\mathbf{u}(t) \in \mathbb{R}^r$, $\mathbf{w}(t) \in \mathbb{R}^n$ are vectors of the input variables, $\mathbf{y}(t) \in \mathbb{R}^m$ is vector of the output variables, $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the nominal system matrix, $\mathbf{A}_d \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times r}$, $\mathbf{B}_w \in \mathbb{R}^{n \times n}$ and $\mathbf{C} \in \mathbb{R}^{m \times n}$ are real matrices.

III. PROBLEM DESCRIPTION

Defining the extended Lyapunov-Krasovskii functional as follows (e.g. see [1], [2])

$$\begin{aligned} v(\mathbf{q}(t), t) = & \mathbf{q}(t)\mathbf{P}\mathbf{q}(t) + \int_{t-\tau(t)}^t \mathbf{q}^T(r)\mathbf{Q}\mathbf{q}(r)dr + \\ & + \int_{t-\tau(t)}^t (\mathbf{y}^T(r)\mathbf{y}(r) - \gamma\mathbf{u}^T(r)\mathbf{u}(r))dr > 0 \end{aligned} \quad (4)$$

$$\mathbf{P} > 0, \quad \mathbf{Q} > 0, \quad \gamma > 0.$$

Problem of the interest is to design asymptotically stable close-loop system with the linear memoryless state feedback controller of the form

$$\mathbf{u}(t) = -\mathbf{K}\mathbf{q}(t) \quad (5)$$

where $\mathbf{K} \in \mathbb{R}^{r \times n}$ is the controller gain matrix.

IV. BASIC PRELIMINARIES

Working with the linear matrix inequalities the concept of symmetric positive definite matrix or positive semi-definite matrix is principally used. The necessary and sufficient conditions for positive definiteness of a symmetric matrix \mathbf{P} are:

- there is a regular matrix \mathbf{Q} , such as that $\mathbf{P} = \mathbf{Q}^T\mathbf{Q}$,
 - all eigenvalues of the matrix \mathbf{P} are positive,
 - all main subdeterminants of the matrix \mathbf{P} are positive
- $$|p_{11}| > 0, \quad \begin{vmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{vmatrix} > 0, \quad \dots, \quad \det(\mathbf{P}) > 0.$$

Schur complement (e.g. see [4], [5]). If $\mathbf{Q} = \mathbf{Q}^T > 0$, $\mathbf{R} = \mathbf{R}^T$, \mathbf{S} are real matrices of appropriate dimensions, then the next inequalities are equivalent

$$\begin{aligned} \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{R} \end{bmatrix} > 0 & \Leftrightarrow \begin{bmatrix} \mathbf{Q} - \mathbf{S}\mathbf{R}^{-1}\mathbf{S}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} > 0 \\ & \Updownarrow \\ \mathbf{Q} - \mathbf{S}\mathbf{R}^{-1}\mathbf{S}^T > 0, \quad \mathbf{R} > 0. \end{aligned} \quad (6)$$

V. STABILITY OF THE SYSTEM

Proposition 5.1: (Lyapunov-Krasovskii Inequality [1], [2]) The linear time delay system of form (1), (2) is stable for $\dot{\tau}(t) \leq \beta$ if exists a symmetric positive definite matrix P and Q and a scalar variable $\gamma > 0$ such that

$$\begin{bmatrix} \Delta_{11} & PA_d & PB & C^T \\ * & -(1-\beta)(Q + C^T C) & 0 & 0 \\ * & * & -\gamma I_r & 0 \\ * & * & * & -I_m \end{bmatrix} < 0 \quad (7)$$

$$P = P^T > 0, \quad Q = Q^T > 0$$

where $\Delta_{11} = A^T P + PA + Q$.

Hereafter, * denotes the symmetric item in a symmetric matrix.

Proof: The result of the functional derivative of $v(q(t), t)$ and inserting (1) into (4) for $u(t) = 0, w(t) = 0, t < 0$ is

$$\begin{aligned} & (Aq(t) + A_d q(t - \tau(t)) + Bu(t))^T P q(t) + \\ & + q^T(t) P (Aq(t) + A_d q(t - \tau(t)) + Bq(t)) + \\ & + q^T(t) Q q(t) - (1 - \dot{\tau}(t)) q^T(t - \tau(t)) Q q(t - \tau(t)) + \\ & + q^T(t) C^T C q(t) - \gamma u^T(t) u(t) - \\ & - q^T(t - \tau(t)) (1 - \dot{\tau}(t)) C^T C q(t - \tau(t)) < 0. \end{aligned} \quad (8)$$

Creating a composite vector $q^*(t)$

$$q^*(t) = [q^T(t) \quad q^T(t - \tau(t)) \quad u^T(t)] \quad (9)$$

the inequality (8) for $\dot{\tau}(t) \leq \beta$ can be rewritten to a matrix form

$$q^*(t)^T \begin{bmatrix} \Delta_{11}^\circ & PA_d & PB \\ * & -(1-\beta)(Q + C^T C) & 0 \\ * & * & -\gamma I_r \end{bmatrix} q^*(t) < 0 \quad (10)$$

where $\Delta_{11}^\circ = A^T P + PA + Q + C^T C$. Using Schur complement property it is obvious that (10) implies (7).

VI. CLOSED LOOP CONTROL

The linear state feedback control law, defined in (5) gives rise to the closed-loop system as follows

$$\dot{q}(t) = (A - BK)q(t) + A_d q(t - \tau(t)) \quad (11)$$

where $A_u = (A - BK)$ is close-loop matrix.

Theorem 6.1: For system (1), (2) the necessary and sufficient condition for the stable nominal control (5) is that there exist positive definite symmetric matrices X, Y and a scalar variable $\gamma > 0$, such that the following LMIs are satisfied

$$\begin{bmatrix} \Delta_{11}^\bullet & X & A_d & B & XC^T \\ * & -Q^{-1} & 0 & 0 & 0 \\ * & * & -(1-\beta)(Q + C^T C) & 0 & 0 \\ * & * & * & -\gamma I_r & 0 \\ * & * & * & * & -I_m \end{bmatrix} < 0$$

$$X > 0, \quad \gamma > 0 \quad (12)$$

where $\Delta_{11}^\bullet = XA^T + AX - BY - YB^T$. The inequality (12) is a bilinear matrix inequalities. Iterative procedure can provide a solution for inequality (12) if the matrix Q is the LMI variable.

Proof: Substituting (11) into (7) gives for $P = P^T > 0$ and $Q = Q^T > 0$.

$$\begin{bmatrix} \Delta_{11}^\circ & PA_d & PB & C^T \\ * & -(1-\beta)(Q + C^T C) & 0 & 0 \\ * & * & -\gamma I_r & 0 \\ * & * & * & -I_m \end{bmatrix} < 0, \quad (13)$$

where $\Delta_{11}^\circ = (A - BK)^T P + P(A - BK) + Q$.

Defining the congruence transform matrix

$$T = T^T = \text{diag} [P^{-1} \quad I \quad I \quad I] \quad (14)$$

and premultiplying (13) from the left and right side by T gives

$$\begin{bmatrix} \Delta_{11}^\triangleright & A_d & B & P^{-1}C^T \\ * & -(1-\beta)(Q + C^T C) & 0 & 0 \\ * & * & -\gamma I_r & 0 \\ * & * & * & -I_m \end{bmatrix} < 0 \quad (15)$$

where $\Delta_{11}^\triangleright = P^{-1}(A - BK)^T + (A - BK)P^{-1} + P^{-1}QP^{-1}$.

Using the Schur complement and substituting

$$P^{-1} = X, \quad Y = KP^{-1} \quad (16)$$

then (15) can be rewritten such that

$$\begin{bmatrix} \Delta_{11}^\triangleleft & X & A_d & B & XC^T \\ * & -Q^{-1} & 0 & 0 & 0 \\ * & * & -(1-\beta)(Q + C^T C) & 0 & 0 \\ * & * & * & -\gamma I_r & 0 \\ * & * & * & * & -I_m \end{bmatrix} < 0 \quad (17)$$

where $\Delta_{11}^\triangleleft = XA^T + AX - BY - YB^T$.

It is evident that (17) implies (12).

VII. ILLUSTRATIVE EXAMPLE

The system is given by (1), (2), where

$$A = \begin{bmatrix} -2.6 & 0.0 & 0.8 \\ -1.2 & 0.2 & 0.0 \\ 0.0 & 0.5 & -3.0 \end{bmatrix}, \quad A_d = \begin{bmatrix} 0.00 & 0.02 & 0.00 \\ 0.00 & 0.00 & -1.00 \\ -0.02 & 0.00 & 0.00 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 0 \\ 7 & 1 \\ 1 & 0 \end{bmatrix}, \quad B_w = \begin{bmatrix} 4 & 0 \\ 7 & 1 \\ 1 & 0 \end{bmatrix}.$$

To solve (12) for LMI matrix variables $X > 0, Z > 0$ and Y we use SeDuMi package for Matlab. The given task was feasible with

$$X = \begin{bmatrix} 0.7458 & -0.1903 & -0.0386 \\ -0.1903 & 0.5741 & -0.1375 \\ -0.0386 & -0.1375 & 0.8125 \end{bmatrix},$$

$$Y = \begin{bmatrix} -0.2375 & 1.5528 & -0.1136 \\ 0.4583 & 0.2085 & 0.4522 \end{bmatrix},$$

$$Q = \begin{bmatrix} 2.4428 & 0.0000 & 0 \\ 0 & 2.4428 & 0 \\ 0 & 0 & 2.4427 \end{bmatrix},$$

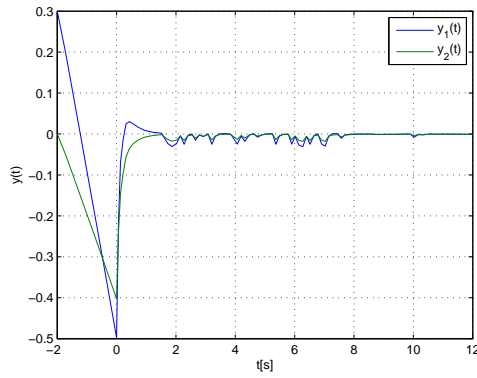
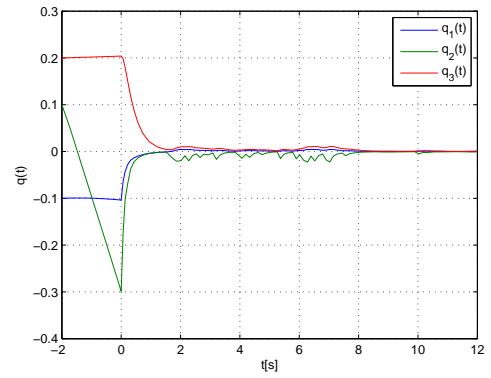
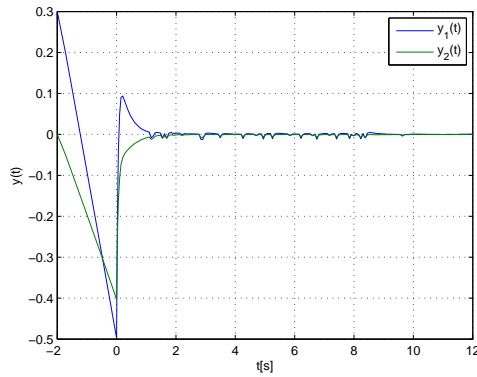
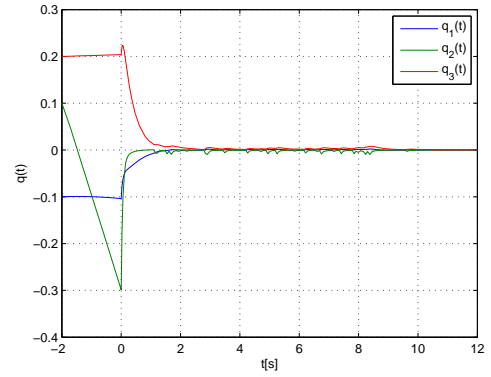
$$Q^{-1} = \begin{bmatrix} 3.8491 & -0.0016 & 0 \\ -0.0016 & 3.8302 & 0.0009 \\ 0 & 0.0009 & 3.8397 \end{bmatrix}, \quad \gamma = 3.9012$$

where $P^{-1} = X$ and $Q^{-1} = Z$. The controller gain matrix K_1 can be obtained from the relation (16) for $\beta = 0.1$, where

$$K_1 = \begin{bmatrix} 0.4528 & 2.9460 & 0.3802 \\ 0.8634 & 0.8259 & 0.7373 \end{bmatrix},$$

$\rho(A_{eu}) = \rho(A + A_d - BK_1) = \{-11.3053, -3.4742, -2.3915\}$

where $A_{eu} = (A + A_d - BK_1)$ is extended closed-loop matrix. It is evident that the eigenvalues spectrum $\rho(A_{eu})$ of the closed control loop is stable.

Fig. 1. Response of the closed-loop system for $\beta = 0.1$ Fig. 3. Response of the closed-loop system for $\beta = 0.1$ Fig. 2. Response of the closed-loop system for $\beta = 0.9$ Fig. 4. Response of the closed-loop system for $\beta = 0.9$

Setting $\beta = 0.9$ the gain matrix \mathbf{K}_2 takes the form

$$\mathbf{K}_2 = \begin{bmatrix} 0.8301 & 6.0715 & 0.3932 \\ 1.0308 & 1.6182 & 0.9657 \end{bmatrix},$$

$$\rho(\mathbf{A}_{eu}) = \rho(\mathbf{A} + \mathbf{A}_d - \mathbf{BK}_2) = \{-21.6199, -3.3515, -2.7161\}$$

and the spectrum of the eigenvalues $\rho(\mathbf{A}_{eu})$ of the closed control loop is stable, too.

Time-delay linear dynamic system (1), (2) is controlled into the steady state with nonzero initial conditions $\mathbf{q}^T(-2) = [-0.1, 0.1, 0.2]^T$ for the above solutions.

In Figure (1) to (4) time response of the closed-loop system with time-varying delay is shown. Solving the problem of stability for the system with time delay is secured only to delay $0 < \dot{\tau}(t) \leq 1$. Simulations were made with the prescribed values $\beta = 0.1$ and $\beta = 0.9$ from above defined interval. Any proposed control response system in the negative time $t \in \langle -\tau(t), 0 \rangle$ is unstable, because one's eigenvalue matrix \mathbf{A}_d is positive. System reaches the steady state under 10.5s. It can be easily seen that the criterion proposed in this paper to design the memory-less feedback controller gives acceptable solutions.

VIII. CONCLUSION

Time delays can appear as a part of the dynamics in many technological processes. Presence of delays in the system is making the analysis and control design more complex because it has a negative impact on the stability of the 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.

The method uses the standard LMI numerical optimization procedures and the iterative procedure to manipulate the system feedback gain matrix as the direct design variable. It is necessary to propose an asymptotically stable closed-loop system. Finally the design method is illustrated by an example.

ACKNOWLEDGMENT

The work presented in this paper was supported by VEGA, Grant Agency of Ministry of Education and Academy of Science of Slovak Republic under Grant No. 1/0256/11. This support is very gratefully acknowledged.

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*. Stanford: 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] NICULESCU, S.I., et al.: *Stability and robust stability of time-delay systems: A quided tour. In Stability and Control of Time-delay Systems*. Springer-Verlag, Berlin, 1998.
- [7] SHAKED, U., et al.: Bounded real criteria for linear time systems with state-delay. *IEEE Transactions on Automatic Control*, 43, 1116-1121, 1998.

Testing the vSphere Storage Appliance in Virtual Environment

Miloš Pavlík (3rd year), Rastislav Hošák (3rd year)

Supervisor: Iveta Zolotová

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

milos.pavlik@tuke.sk, rastislav.hosak@tuke.sk, iveta.zolotova@tuke.sk

Abstract—this article deals with testing of VMware VSA (vSphere Storage Appliance), which combines storage directly attached to two or three physical servers and combines it into a single (iSCSI) storage. The first point of article is to test if this block-access storage pool is accessible by applications that are and are not running in virtual machines (VMs) which are running in these servers. Another point of this article is to discover the possibilities of VMware VSA solution.

Keywords—vSphere storage appliance, ESXi, storage, cluster.

I. INTRODUCTION

As a part of our research and implementing of VMware software on Dept. of Cybernetics and Artificial Intelligence, we've decided, that we'll try new VMware VSA appliance. We had problems with getting the NAS (Network Attached Storage) or SAN (Storage Area Network) storages, so we decided to use this appliance. In fact, that should to solve our problems. We are using virtual iSCSI (Internet Small Computer System Interface) network storage through *Openfiler 2.99* software, but this solution seems not to be the best way how create shared data storage for parts of our virtual architecture (vSphere platform, High Availability cluster, Fault Tolerance).

II. VSPHERE STORAGE APPLIANCE

A. VSA

vSphere Storage Appliance runs on multiple servers simultaneously, so you can have the confidence that data is available to any of your workloads, even when a server fails. With vSphere's availability features, we can restart virtual machines automatically when a server fails, use fault tolerant protection so applications run uninterrupted, and perform maintenance with zero impact to the end user by migrating applications live from one server to another using vMotion. [1, 4].

With the vSphere 5 came the VSA appliance, which basically combines storage directly attached to two or three physical servers and combines it into a single (iSCSI) storage pool. It is a block-access pool accessible by applications running in virtual machines (VMs) in these servers. The appliance uses the available space on the local disk(s) of the ESXi servers and present one replicated NFS (Network File

System) volume per ESXi server. Created storage pool is built by using the VSA software, which is running inside individual virtual machines (VMS) in the servers. In fact, it is a way how to provide shared storages without any need to purchase a physical SAN or NAS. It use the local storage in each ESXi server, and present it to the VSA virtual machine, and that in turn will present it back to the ESXi servers as shared storage. Each ESXi server has a VSA virtual machine running, and each will contain a duplicate of the other's storage (figure 1). VMware VSA supports the Storage vMotion movement of apps between servers to shared storage with no disruption of service access. VSA is managed through vCenter Server and the installation process handles network set up and vSphere HA (High Availability) implementation. It also handles the installation and configuration of vCenter Server, vSphere as well as VMware VSA itself [4, 5].

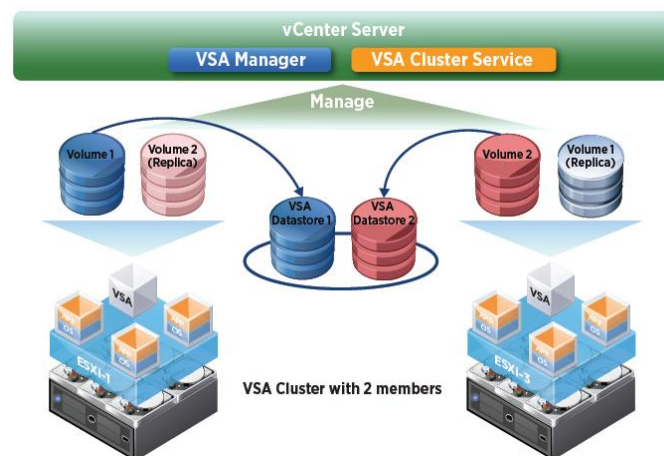


Fig. 1 vSphere Storage Appliance, two nodes.

B. Requirements and basic facts

First of all there is needed to make clear in some facts. The only supported storage configuration is RAID 10. That means that the VSA virtual machines are using RAID 1 configuration for purposes of High Availability cluster. That brings a question. Is it effective? Well, not only we came to a conclusion, that RAID 10 configuration is highly ineffective to our local storages. We have 4 x 60 GB drives

in each our ESXi server. After setting up a RAID 10 configuration on each ESXi server the total size of local server storage is 120GB, which is the half of size. After that, the VSA reserves another 50% for VSAs duplicate. So after summarize we have 25% of our initial storage space.

Requirements:

- vSphere Essentials Plus, Standard, Enterprise, or Enterprise Plus
- VMware vCenter Server
- Two to three ESXi hosts (Fresh installs only)
 - 4 NICs (Network Interface Card) on each host
 - Local Storage in Raid 10 config.
 - NO configuration changes

The vCenter Server can be physical or virtual, but if the vCenter is in virtual machine it must not run on any of ESXi hosts that are hosting the VSA appliance. In other words they must not be within the same Cluster as VSA appliances. Four NICs, mentioned in requirements, are divided to two groups. First one, with two NICs is used for backend of the VSAs, replication the duplicates. Backend is basically is a class of storage interface. The second group with another two NICs is used for synchronization.

III. INSTALLING & TESTING THE VSPHERE VIRTUAL STORAGE APPLIANCE

At the beginning and before the testing VSA, we had to be sure, that we meet the requirements. The biggest problem for us was the requirement of 4 NICs. Our IBM x3650 M2 physical hosts have only two of them, so it was impossible to install VSA but then we decided that we can use virtual environment. So we had to install our virtual environment. All necessary VMs are installed on host with IP: 147.232.61.71. First step was to add a networking hardware to host, virtual machine, which handle the traffic for this connection and its network label is Trunk. Next step was adding a new resource pool called VSA Cluster, which will represent our testing virtual cluster. After that we have installed two vESXi (virtual ESXi) servers, because we wanted the two node solution. Each of them has four NICs, one is VM network and three others are our created Trunk. Each of vESXi servers has four 135 GB hard drive (virtual). Now when we have installed our vESXi servers, we've created another datacenter called Test and we've created a new VSA cluster in it. For this cluster are used two created virtual vESXi servers [3, 6].

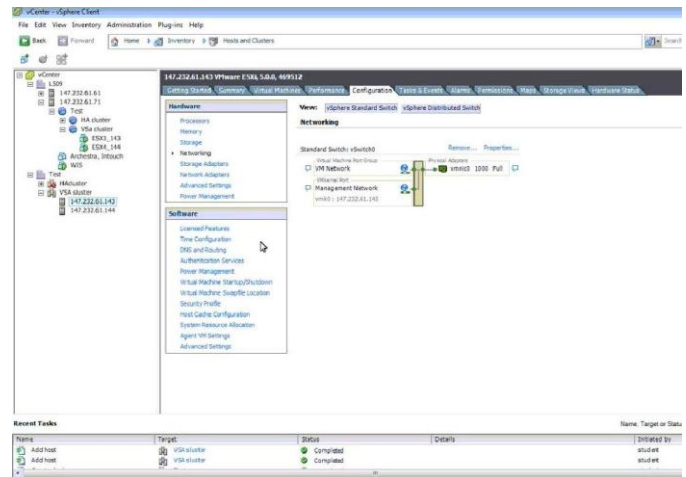


Fig. 2 Configuring VSA cluster in VMware vCenter Server

Now we could install VSA. The installation had a several steps. First is the review of vSphere features that the vSphere Storage Appliance enables. Next one is choosing datacenter (VSA cluster). In another step we had to select our vESXi servers we wanted to use to host a VSA VM. Of course the host must have processors that family is the same. Fourth step was managing the VSA cluster network, so the VSA cluster IP, IP for each Host, had to be set. Once the first IP is entered, it will automatically fill in sequential IPs in the following fields. Now after formatting of storage the installer configures VSA cluster network, deploys VSA, installs VSA and mount the data stores [7].

At this point, we have 2 vESXi servers, now in a Cluster with vMotion and HA enabled. Each host already has LUNs (Logical Unit Number) presented to it from each VSA via NFS, and is ready for use.

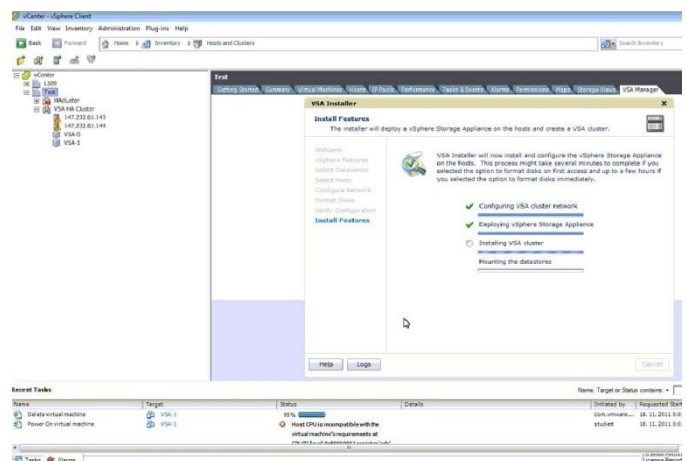


Fig. 3 Installing and configuring VSA

After testing we decided that each vESXi server has a VSA deployed to it as a Virtual Machine (VSA0, VSA1). The appliances use the available space on the local disk(s) of the vESXi servers and present one replicated NFS volume per vESXi server. The NFS data stores exported from the VSA can now be used as shared storage on all of the vESXi servers in the same datacenter. The VSA creates shared storage out of local storage for use of a specific set of hosts.

IV. CONCLUSION

Testing of vSphere Storage Appliance was possibility to solve problem with shared data storage, because it is a way how to provide shared storages without any need to purchase a physical SAN or NAS. It combines storages directly attached to two or physical servers and combines it into a single (iSCSI) storage pool. The VSA itself is interesting idea but it has many interesting shortcomings as well. First of all is the fact that the only supported storage configuration is RAID 10, which is in fact very ineffective, because of 75% of local storage capacity loss, as was mentioned above. We wanted to try, if applications in VMs are executing in servers that are not part of the vSphere Storage Appliance, if they could access the VSA and if it isn't a virtual NFS filer only. And we have confirmed information taken from VMware sources that applications are not accessing that storage. This is not intended to be a NFS filer, but instead a store for VMs. So the VMs are stored on the VSA and data is stored inside the VMs, not directly on the storage.

ACKNOWLEDGMENT

The work presented in this paper was supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic within the KEGA 021TUKE-4/2012 project (70%) and VEGA - 1/0286/11 project (30%). This work was concluded in our laboratory with the software and licenses provided by VMware company.

REFERENCES

- [1] L. Landryová, I. Zolotová, "Challenges and Software Aspects of Remote Labs for Engineering Education." in Proceedings of 8th International Conference on Information Technology Based Higher Education and Training. Kumamoto, Japan, available online at URL: <http://ithet07.coe.kumamoto-u.ac.jp>, July 10-13.2007, pp 1-4.
- [2] 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.
- [3] FRANEKOVÁ, M.- KÁLLAY, F.- PENIAK, P. VESTENICKÝ, P.: Safety Communication of Industrial Networks. (In Slovak), EDIS, ŽU Žilina, 2007, ISBN 978-80-8070-715-6, monografia
- [4] <http://www.vmware.com/products/datacenter-virtualization/vsphere/vsphere-storage-appliance/overview.html>
- [5] <http://blogs.vmware.com/vsphere/2011/08/new-enhanced-vsphere-50-storage-features-part-4-vsphere-storage-appliance.html>
- [6] <http://www.vmware.com/products/datacenter-virtualization/vsphere/vsphere-storage-appliance/overview.html>
- [7] <http://blogs.vmware.com/vsphere/2011/08/new-enhanced-vsphere-50-storage-features-part-4-vsphere-storage-appliance.html>

Three-Dimensional Software Visualization and UML

¹Kristián ŠESTÁK (4st year)
Supervisor: ²Zdeněk HAVLICE

¹Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

²Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹kristian.sestak@gmail.com, ²Zdenek.Havlice@tuke.sk

Abstract— this paper describes visualization technique, types and visualization of the static aspects in the software system. Graphical visualization of software has the potential to result in a better and faster understanding of its design and functionality. We present an overview of current research in the area. In this paper we survey visualization techniques both 2D and but mainly in 3D-based representing the static aspects of the software and its evolution.

Keywords—software comprehension, visual representation, software visualization, static aspects visualization, visualization in 3D, UML, graphical representation

I. INTRODUCTION

Conventional programming languages are difficult to learn and use, requiring skills that many people do not have. One approach to the problems of comprehension and complexity is through computer graphics and visualization. The essential idea is that visual representations can make the process of understanding software easier.

This is software visualization, which encompasses all applications of graphics to software development, including static program visualization, visual programming and visualization of run-time behavior.

Also help software engineers cope with complexity and to increase programmer productivity is through visualization. Experiments have shown that using visualization techniques in a software development project increases the odds of succeeding.

We focus on research in the areas of Knowledge-based software life cycle and architectures. [1], [2], [3], [13]

II. UML

Unified Modeling Language (UML), as described by Booch, Rumbaugh et al. [1998], “is a graphical standard for visualizing, specifying, constructing, and documenting the artefacts of a software-intensive system”. The UML has gained wide acceptance in the software engineering.

UML does not limit the use of 3D to diagrams. Stereotypes allow the creation of new visualizations which give greater meaning than a standard type.

UML provides distinct diagrams for static and dynamic information. The static visualizations of a system, (e.g., UML class diagrams), describe the static structure of a system in terms of software entities and their relationships. To represent

dynamic aspects of a system, UML provides object diagrams and sequence diagrams. Their main problem is, that they do not scale beyond a certain number of events because of physical constraints, which can however be partially overcome with the help of interactive exploration techniques

The UML can help even non-programmers get a better grasp of the overall functionality of the system. [4], [5], [8]

III. SOFTWARE VISUALIZATION

Software visualization is a discipline that makes use of various forms of imagery to provide insight and understanding and to reduce complexity of the existing system under consideration.

The basic motivation is program understanding. We want to use visualization to understand our designs, our code the structure of our code, our data structures, and the execution of the code the history of the code and the uses of the code. Most of today visualization systems claim to do some form of understanding.

A definition of software visualization suggested by Diehl and another for program visualization defined by Price will be used. [16],[17]

Stephan Diehl defines software visualization as “the art and science of generating visual representations of various aspects of software and its development process”. [14]

Card et al. define visualization as “the use of computer supported, interactive, visual representations of data to amplify cognition”, where cognition is the acquisition or use of knowledge. [15]

Software visualization is a wide research area, involving a large number of visualization terms and techniques. In the past, a number of taxonomies have been proposed to categorize and classify software visualization terms, types and techniques.

Software visualization can be seen as a specialized subset of information visualization, because software visualization is the process of creating graphical representations of abstract, generally non-numerical data, while information visualization refers to visualizing numerical as well as non-numerical data.

Software visualization allows developers to discover quality defects automatically during the software development life cycle and is particularly applied to legacy code maintenance where maintainers may not have previously seen the system.

Based on a number of resources, especially Kang Zhang’s papers, a model for software visualization fields is introduced,

as shown in Fig. 1. [6], [7], [9], [10], [11], [13]

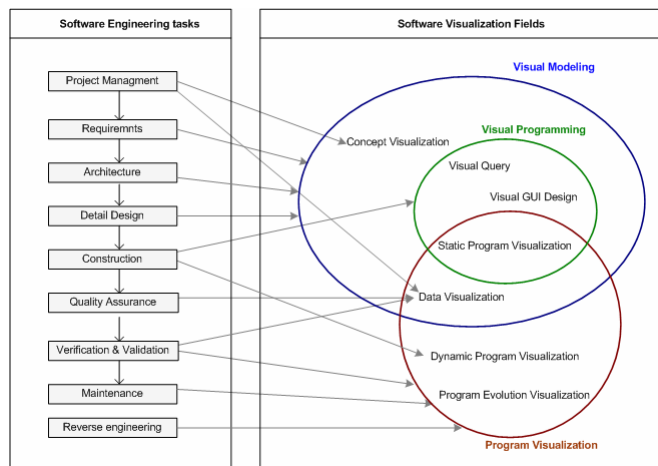


Fig. 1 Software visualization fields used by Software Engineering tasks

A. Software visualization types

A set of visualization techniques that share fundamental display mode features can be organized under a specific group name. This group is called the visualization type.

- Montion
 - Static
 - Dynamic
- Dimension
 - 2-dimensional
 - 3-dimensional
- Construction
 - Nodes and links
 - Space filling
 - Matrix
 - Chart
 - Text based
 - Shape based
 - Combinations

B. Software visualization techniques

Visualization techniques are forms of visualizations consisting of a collection of elements such as points, lines, shapes, texts, and textures. Each of these elements represents an entity or an attribute from the dataset that needs to be visualized. Each visualization technique belongs to a visualization type.

Every program visualization technique is visually static (motionless) or dynamic (be in motion). Static visualization techniques, such as architecture diagram, display the static aspects of software, while the dynamic visualizations such as algorithm animations visualize the dynamic behaviors of software.

Most of the 3-dimensional program visualizations are the optimizations techniques of 2-dimensional version of the same visualization technique.

Visualization technique elements are visualizable parameters that affect viewer comprehension such as shape, size, color and texture.

Any visualization technique is composed of a set of these basic elements that represent the software information. Selecting the appropriate set of these elements is essential for the success of the visualization because some of these

elements are more effective in showing the differences between the visualization components, while others are more effective in displaying the properties of visualization components. [6]

IV. VISUALIZATION OF THE STATIC ASPECTS

The visualization of the static aspects of software focuses on visualizing software as it is coded, and dealing with information that is valid for all possible executions of the software. [12]

A. Visualizing Software Organization

Panas et al. [18] developed a very detailed Cities metaphor to represent software. Their metaphor is as close as possible to real cities, with an extremely detailed and realistic visualization, with trees, streets, street lamps (Fig. 2). [12]

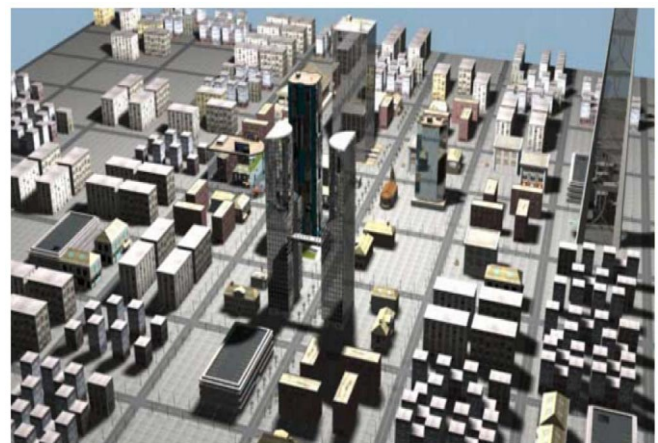


Fig. 2 Realistic City metaphor representing software from

Graham et al. [19] proposed another real-world metaphor to represent software: the Solar system. This visualization represents the software as a virtual galaxy made up of many solar systems (Fig. 3).

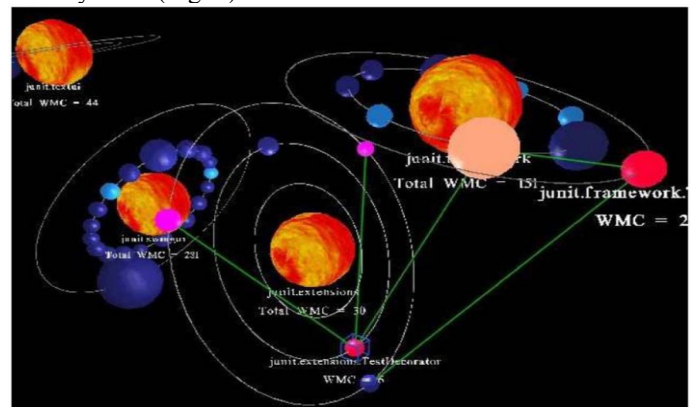


Fig. 3 Solar System metaphor from

B. Visualizing Relationships in the Software

Visualizing relationships in the software is a harder task than visualizing the software hierarchy, because components can have a much larger number of relations of many kinds, such as inheritance, method calls, and accesses

In Figs. 4a and 4b, we can see a UML class diagram and its equivalent geon diagram. Any UML class diagram can be turned into its equivalent geon diagram [12].

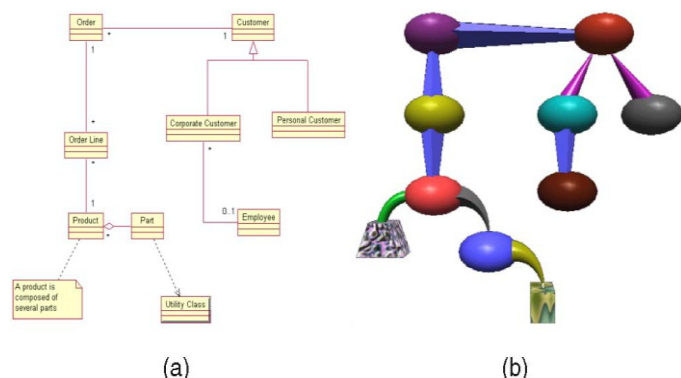


Fig. 4 (a) UML diagram. (b) Equivalent geon diagram

Alam and Dugerdil [20], [21] proposed another way to visualize relationships with a City metaphor. Their visualization technique named Evospaces displays relationships as solid curved pipes between buildings (classes) or between objects inside buildings (methods). A colored segment moving along a pipe suggests the direction of the relation from the origin to the destination (Fig. 5). [12]

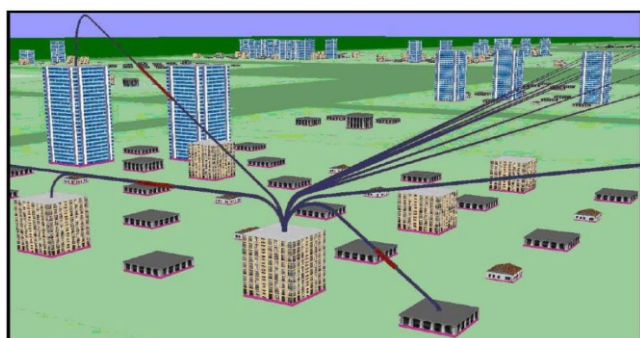


Fig. 5 Evospaces visualization

Panas et al. [25] use a Cities metaphor to represent software. We name it “The Unified Single-View City.” In their visualization (Fig. 6), methods are represented by buildings, which are placed on blue plates that symbolize classes. [12]

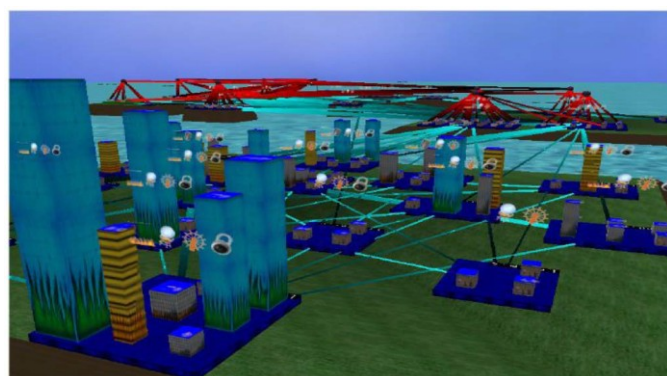


Fig. 6 The Cities and Island metaphor from

Soon after the Software Landscape visualization, Balzer and Deussen [23] developed another 3D visualization also based on a clustered graph layout, to display large and complex graphs (Fig. 7). Their technique uses clustering, dynamic transparency, and edge bundling to visualize a graph without altering its structure or layout. The main idea is to group remote vertices and classes into clusters. [12]

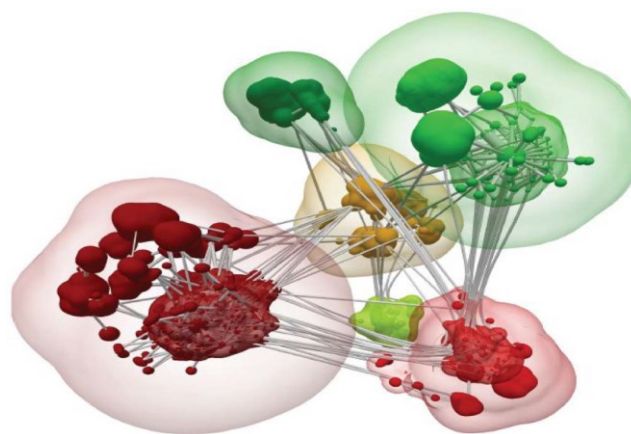


Fig. 7 Clustered graph layout

C. Metric-Centered Visualization

Software metric is a numeric measure of some property of a piece of software or its specifications. Software metrics are interesting because they provide information about the quality of the software design

Polymetric views, created by Lanza et al. [26], is a lightweight visualization of software enriched with software metrics (Fig. 8).

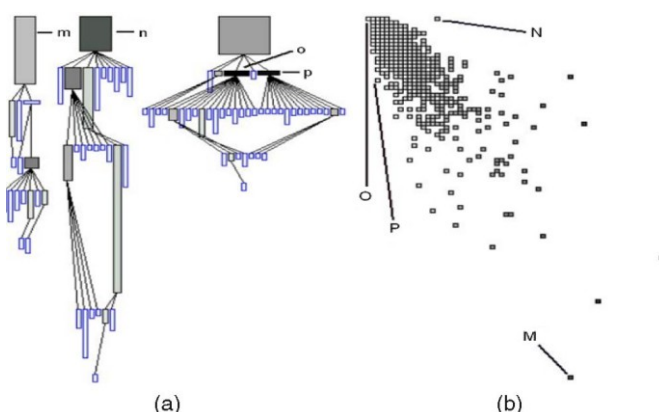


Fig. 8 Polymetric views from [36]: (a) Inheritance Tree. (b) Correlation graph

Langelier et al. [22] think that the simplicity of the chosen visual representation is crucial for human perception. Their visualization, named “VERSO,” uses simple boxes to represent classes, and cylinders for interfaces (Fig. 9). The human visual system is a pattern seeker of enormous power and subtlety, and relying on these simple shapes fosters a better and faster recognition of the underlying information without overloading the visualization. [12]

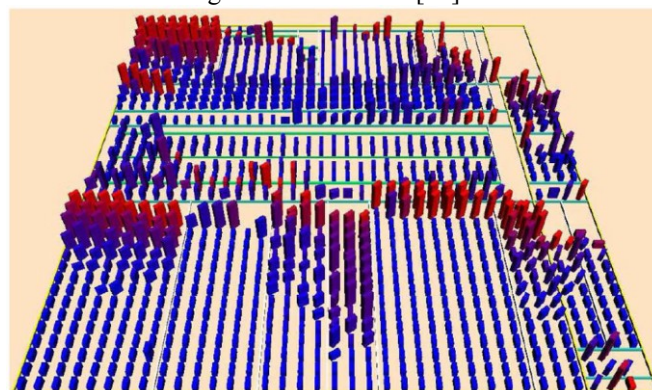


Fig. 9 VERSO

CodeCity is a visualization tool proposed by Wettel and Lanza, [24] that represents the software with a City metaphor. This visualization is also available as a stand-alone tool and as an open-source 3D software visualization plug-in for Eclipse. CodeCity displays classes as buildings, and packages as city districts. Classes within the same package are grouped together, using a Treemap layout to slice the city into districts (Fig. 10). [12]

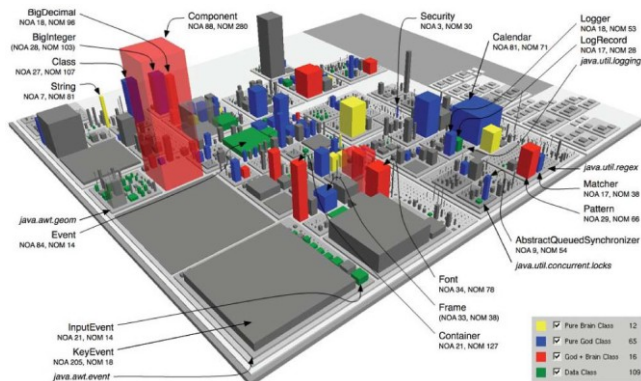


Fig. 10 CodeCity class-level disharmonies map

V. CONCLUSION

Software developers want to do their development. They want to do it as quickly, as accurately, and as high quality as possible. They will use whatever means they can to achieve these goals.

A trend of this decade is to experiment with new 3D visualizations. Promising results have been obtained even though 3D navigation and interactions are problematic. Computer-based visualization and software should be an ideal match. Visualization should be central to our programming experience. Programmers should be using it when as they develop programs, when they analyze systems statically, when they look at the dynamic behavior of their systems when they are debugging when they are testing, and even when they are editing or modifying existing code.

But programmers quickly give up the UML for header or interface files and ignore the UML notations once the program is written. Visual languages work well in specific domains but, despite much work, have had little success for general programming.

We first have to comprehend what are the understanding problems that software visualization should be used to address. The third reality that software visualization needs to address to be successful is the reality of how developers work and what they need to do their work. Despite the benefits that many papers claim, software visualization is not yet a fully accepted part of the software engineering process in the software industry. Another important problem of many existing software visualization methods and tools is their limited scalability with respect to the huge size of modern software systems and different datasets that need to be visualized [7], [12]

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0008-10 and by the Scientific Grant Agency of the Ministry of Education - VEGA Grant No. 1/0531/12 Use of Knowledge in

Software Processes.

REFERENCES

- [1] Calum A. McK. Grant, Software visualization in Prolog Issue 511 of Technical report (University of Cambridge. Computer Laboratory), University of Cambridge, Computer Laboratory, 2001
- [2] Thomas Ball and Eric E. Sumner, Software visualization in the large, Computer, 1996.
- [3] Brad A Myers, Taxonomies of Visual Programming and Program Visualization, Volume: 1, Issue: 1, Publisher: Elsevier, Pages: 97-123
- [4] McIntosh, P., Hamilton, M., Ron van Schyndel, X3D-UML: Enabling Advanced UML Visualisation Through X3D, School of Computer Science and Information Technology RMIT University GPO Box 2476V, Melbourne, VIC., Australia, 2005
- [5] SHARP, R., AND ROUNTEV, A. 2005. Interactive exploration of uml sequence diagrams. In Proceedings of VISSOFT 2005 (3rd IEEE Workshop on Visualizing Software for Understanding and Analysis, IEEE CS Press, 8–13.
- [6] Qais Ali, Static Program Visualization within the ASF+SDF Meta-Environment, Organisation: Centrum voor Wiskunde en Informatica, 2008
- [7] Steven P. Reiss, The Paradox of Software Visualization, 3rd IEEE International Workshop on Visualizing Software for Understanding and Analysis. VISSOFT 2005, pages 1-5, 2005
- [8] Benjamin N. Hoipkemie, Nicholas A. Kraft, Brian A. Malloy, 3d visualization of class template diagrams for deployed open source applications (2006)
- [9] Teyseyre, A., and Campo, M., An Overview of 3D Software Visualization, 2008
- [10] S. Card, J. MacKinlay, and B. Shneiderman, Eds., Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann Publishers, 1998.
- [11] Petre, M., Ed de Quincey, A gentle overview of software visualisation Centre for Research in Computing, The Open University, Walton Hall, Milton Keynes, UK and School of Computing and Mathematics, Keele University, UK., 2006
- [12] Caserta, P., Zendra, O., Visualization of the Static Aspects of Software: A Survey, TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 17, NO. 7, JULY 2011
- [13] Robitaille, S., Schauer, R., Keller, R.K., Bridging Program Comprehension Tools by Design Navigation, In Proceedings of IEEE International Conference on Software Maintenance (ICSM 2000), S. 22-32, 2000.
- [14] Storey, M.-A., Theories, Methods and Tools in Program Comprehension, Program Comprehension, 2005. IWPC 2005. Proceedings 13th International Workshop on volume, Issue, 15-16 May 2005, pages 181 – 191, 2005
- [15] Card, S., Mackinlay, J., and Shneiderman, B. Readings in information visualization: using vision to think. Morgan Kaufmann, 1999.
- [16] Price, Baecker and Small, A Taxonomy of Software Visualization, Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences, vol 2, pages 597-606, 1992
- [17] Diehl, Stephan, Software Visualization: Visualizing the Structure, Behaviour, and Evolution of Software, ISBN 3540465049, XII, Springer, 2007
- [18] T. Panas, R. Berrigan, and J. Grundy. A 3d metaphor for software production visualization. International Conference on Information Visualization, page 314, 2003.
- [19] H. Graham, H. Y. Yang, and R. Berrigan. A solar system metaphor for 3D visualisation of object oriented software metrics. In Proc. Australasian Symposium on Information Visualisation, pages 53–59, 2004
- [20] S. Alam and D. Ph, "EvoSpaces: 3D Visualization of Software Architecture," Proc. Int'l Conf. Software Eng. and Knowledge Eng., 2007.
- [21] S. Alam and P. Dugerdil, "EvoSpaces Visualization Tool: Exploring Software Architecture in 3D," Proc. 14th Conf. Reverse Eng., pp. 269-270, 2007.
- [22] G. Langelier, H. Sahraoui, and P. Poulin, "Visualization-Based Analysis of Quality for Large-Scale Software Systems," Proc. 20th IEEE/ACM Int'l Conf. Automated Software Eng., pp. 214-223, 2005.
- [23] M. Balzer and O. Deussen, "Level-of-Detail Visualization of Clustered Graph Layouts," Proc. Asia-Pacific Symp. Visualization, 2007.
- [24] R. Wettel and M. Lanza, "Visualizing Software Systems as Cities," Proc. Fourth IEEE Int'l Workshop Visualizing Software for Understanding and Analysis, 2007.

- [25] Panas, T., Berrigan, R., and Grundy, J.. A 3d metaphor for software production visualization. In IV '03: Proceedings of the Seventh International Conference on Information Visualization, page 314, Washington, DC, USA, 2003. IEEE Computer Society.
- [26] Lanza, M. and Ducasse, S., Polymetric Views — A Lightweight Visual Approach to Reverse Engineering. IEEE Transactions on Software Engineering, vol. 29, no. 9, pages 782–795, September 2003.

Usage of Hybrid Honeypot as an Intrusion Detection System Mechanism

¹Peter FANFARA (2nd year), ²Adrián PEKÁR (1st year)
Supervisor: ³Liberios VOKOROKOS

^{1,2,3} Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

¹peter.fanfara@tuke.sk, ²adrian.pekar@tuke.sk, ³liberios.vokorokos@tuke.sk

Abstract—System security is one of many informatics sections that have massive progress. Each system works with some type of sensitive information and that's the reason why is Intrusion Detection System widely used in system security. Security in these systems is alpha and omega. Main IDS disadvantage represents the weakness of detection mechanism. Paper attention is mainly devoted to advanced technology called Honeypot and usage of this technology as a detection mechanism of IDS.

Keywords—Honeypot, Intrusion, Intrusion Detection System, Types of Honeypots.

I. INTRODUCTION

Due to the rapid spread of Internet and Web technologies people can quickly and easily find information and send messages. However, if we don't put at the same time a sufficiently high priority to the basic systems security, hackers can take over systems using malicious code through existing system vulnerabilities and program weaknesses. After the attacker's invasion, destruction, theft and falsification of information will cause major damage to most of companies and personal property. These days, resulting from potential threats, arises a growing interest in improving information security as well as intrusion detection.

Beginnings of detection intrusions have brought the complications. Between the theoretical and practical plane of intrusion detection still exists gap. Well established defense of network/system is based on using firewall and intrusion detection system (IDS). Once the attackers are aware that firewall has allowed an exception for external security service, they are able to use this service to gain access to internal servers through firewall and make another attack. IDS cannot provide additional information for detection of enemy attacks and cannot reduce losses caused by those attacks.

Conventional approach to security is considerably gently focused on defense, but interest is increasingly devoted to more aggressive defense forms against potential attackers and intruders. Example of this form is also protection against intrusions based on the bait through Honeypot [6].

The following chapters describe system security using IDS with detection mechanism based on advanced Honeypot technology.

II. INTRUSION DETECTION SYSTEM

IDS can be defined as a tool or software application that monitors activities of computer system and/or network, due to potential occurrence of malicious activities or breaching security policy. IDS produce reports for control station. It's primarily focused on identifying and recording information about any events as well as reporting similar attempts [10].

A. Classification of Intrusion Detection System

In view to the various application environments the IDS can be classified into two general types [1]:

- *Host-based* – consists of an agent located on host computer which is used for continuous monitoring information from system audit data or network activities logs. This IDS sensor type typically includes a software agent. If there are unusual circumstances, system automatically generates and sends a warning.
- *Network-based* – independent platform for intrusions identification using direct capturing of transmitted network packets and monitoring several computers. Detection sensors are placed in network bottlenecks – capture all network traffic and analyze individual packets content looking for dangerous operation.

Given the detection method IDS can be divided into next three types [1]:

- *Anomaly detection* – refers to the pattern found in the data set that is not consistent with normal behavior. Anomaly detection provides basic performance for normal network traffic. The alarm sounds only if the current network traffic is beyond basic parameters.
- *Misuse detection* – collects previous hacker attack characteristics and patterns which are then saved to knowledge attack database. Consequently it can identify attacks with same patterns and characteristics as previously stored attack. IDS cannot trigger alarm if hacker uses new attack method that hasn't been previously reported or detected.
- *Hybrid mode detection* – represents attack detection using previous two types which resulting in reduction of generating false alarms.

B. IDS structure and architecture

IDS consist of several elements where the main element is a sensor – mechanism for analysis, responsible for intrusion detection. This sensor contains a mechanism that makes

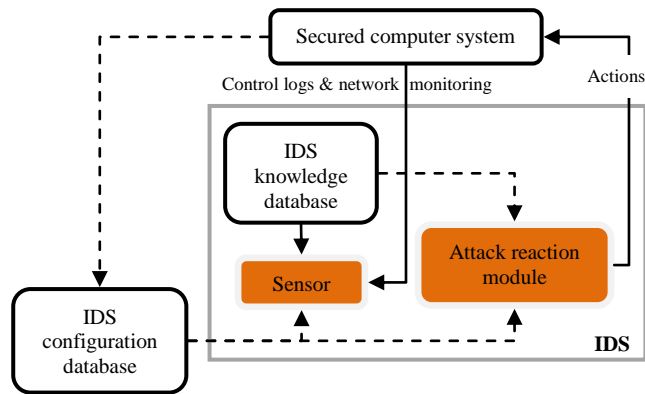


Fig. 1 Intrusion detection system structure

decisions regarding breach. Sensor receives data from three main sources of information (Fig. 1): IDS knowledge database, system logs and audit trails. System logs may include e.g. file system configuration and user permissions. This information forms the basis for further decision on intrusion detection.

Sensor is integrated together with element responsible for data collection (Fig. 2) – events generator. The data collecting method is set by policy of events generator that defines filtering method for events information notifications. Events generator (operating system, network & application) in accordance with security policies produced set of events (system logs, control records or network packets). These occurrences together with information policy may be stored either in a protected system or outside it. In some cases they're not stored, e.g. when events streams are directly transmitted to analyzer, especially network packets [8].

The role of sensor is filtering information and discarding any irrelevant data obtained from event file related to protected system and detecting suspicious activity. For this purpose sensor uses detection policy database, which is composed from following portions: pattern attack, normal behavior, profiles and necessary parameters. Database contains IDS configuration parameters & communication method with reaction module. Sensor has custom database that also includes a dynamic history of potential intrusions [8].

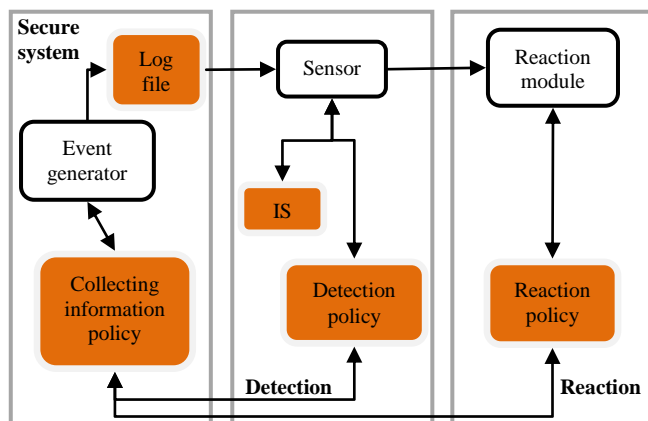


Fig. 2 Intrusion detection system elements

C. Intrusion detection tools

Nowadays exist many IDS and may be specific to system by using their own tools. Mostly used tool is Snort which in addition has excellent conditions for use to enhance distributed system security in combination with Honeypots.

Snort represents open-source IDS that can detect and warn of attack (e.g. against Honeypot). It can also capture packets and network load given by packets included in the attack. Collected information may be critical in analyzing attacker's activities. For designing is Snort using modular architecture and rules based language. It combines abnormal behavior, detection signature and different protocol detection methods [11].

In order to be able the most successfully monitor hackers activities in distributed computer systems, it was domesticated the methodology of lying and cheating by providing emulation of some system services. This system at first sight appears to be legitimate. Due to penetration and clarification of various attackers tactics is possible to record and monitor all hackers' activities. This idea is assumed by using an advanced security tool called Honeypot.

III. HONEYPOT

Honeypot is closely monitored network decoy available in different shapes and sizes, serving for various purposes. It can be placed in a computer network with firewall – before or/and behind firewall – these are the most frequented attackers sites for obtaining access to system. These sites provide the best solution to get maximum information about attackers' activities. The aim of Honeypot is to acquire information with compromising system data in a way that in the future will be any system infiltration unfeasible.

The main Honeypot benefit is in detection. With its simplicity can address IDS shortcomings – minimizing amount of generated false positive and false negative alerts. There are several situations in which IDS cannot generate warning on attack: attack is too short, appropriate security rule refers too many false alarms or detects excessively network traffic and thus dropping packets. One solution is to use Honeypot, because it has no way affect onto system function. Honeypot implementation use unused IP address, which means all incoming communication is almost certainly unauthorized – i.e. no false positive or false negative alarm warning or large data files to analyze [5].

Data obtained from Honeypots can be used to create better protection and countermeasures or reconfiguration against future threats.

A. Types of Honeypots

Honeypots can be classified in different ways. Most frequently used is classification according to purpose and level of interaction.

Purpose of Honeypots

This basic classification divides Honeypots based on area of deployment.

- **Research Honeypot** – this type is used purely for research. Main objective is to get as much information as possible about intruder in a way that allows fully infiltration and penetration of security system. It's used to obtain information and detect new methods and types of tools used to attack other system as well as analyzing hacker's traces, their identity or modus operandi. Another option of research Honeypot is to use it in discovering potential risks and information vulnerabilities in enterprise system [2].

Primary function is to examine how attackers proceed and lead attacks – understanding their motives, behavior

& organization. Research Honeypots are complex in terms of deployment, maintenance and capturing huge amounts of data. On the other hand they are highly-useful security tool in field of development and forensic analysis capabilities.

Despite the obtained information from research Honeypot that can be used to improve prevention against attack. Improve detection and response to attack this Honeypot type contributes totally to direct security only in small hand [3].

- Production Honeypot – it's used in organization for protection and help in reducing level of risk – provides immediate enhancing system security [4]. Because it doesn't require such functionality as research Honeypot its development and deployment is usually much easier. Although it can identify various attack methods, production Honeypot provides less information about attacker than research Honeypot. It's possible to determine from where attackers come from and what specific action performed, but it cannot determine intruders' identities, how they are organized or which tools were used.

Production Honeypot has minimum value as a prevention mechanism. The best way of implementing Honeypot is to use well firewalled system, IDS, mechanism for locking and fixing system [3].

Level of interaction

All Honeypots work on the same concept – nobody should interact with Honeypot. Level of interaction can be defined as maximum range of options available to attack allowed by Honeypot. Therefore any transactions or interaction based on definition become illegitimate. Honeypots can be also categorized according to level of interaction between intruders and system. This classification helps in choosing correct type for deploying in system [3].

- Low-interaction – doesn't contain any operating system (OS) for communication with attacker. All tools are installed purely for emulation of OS and services that cannot be used to gain full access to Honeypot. Emulation is set up to cooperate with attacker and malicious code resulting in radical risk reduction. Attackers can only scan Honeypot and connect to several ports. Low-interaction Honeypots are characterized by possibility of easy deployment and maintenance. Low-interaction Honeypot example is Honeyd.
- Medium-interaction – this type is more sophisticated than previous, but still doesn't have installed OS. Medium-interaction Honeypot only provides to attacker an illusion of OS, because it contains a number of emulated services, with which can interact. This type can detect automated attacks and extract information about getting malware binaries – malicious software can be automatically downloaded and analyzed. Some example of this Honeypot type includes Dionaea and Honeytrap.
- High-interaction – the most advanced Honeypot. On the other hand it represents the most complex and time-consuming design with the highest rate risk, because it implies functional OS. It gives attacker ability to communicate with real OS where is nothing simulated, emulated or restricted. This Honeypot allows collect the highest amount of information as it can detect and analyze all performed activities. Main accent is set to

obtain valuable information about intruders by making available entire system or even allowing the handling of.

B. Architecture of hybrid Honeypot

Hybrid Honeypot represents combination of two Honeypots with different levels of interaction. Combination is safe solution, because it's possible to take advantages of both Honeypots types that complement each other and thus limit their disadvantages (TABLE I). Ideal solution is using a low-interaction Honeypot with high-interaction. Low-interaction Honeypot acts as a lightweight proxy which relieves high-interaction Honeypot and allows focusing on processing all IP address space network traffic [4].

TABLE I
THE ESSENCE OF HYBRID HONEYPOT

Low-interaction Honeypot	High-interaction Honeypot	Hybrid Honeypot
+ fast	- slow	+ fast
- no possibility to detect unknown attack	+ possibility to detect unknown attack + 0 false produced warnings	+ possibility to detect unknown attack + 0 false produced warnings
+ resists to time-bomb + handles interaction with attackers	- unable to resist time-bomb and can't handle interaction with attackers	+ resists to time-bomb + handles interaction with attackers
+ cheap	- expensive	+ relatively expensive
+ simple to set up and maintain	- complicate to set up and maintain	- complicate to set up and maintain

For each Honeypot proposal is impossible to not use implementation tools, which have considerable importance in improving systems security.

Dionaea is modular architecture using low-interaction Honeypot. It's capable to simulate server main services and vulnerabilities, because of attracting attacker/attack attention or withdrawal malicious code [12].

Sebek is the most advanced tool for comprehensive data collection, aiming to capture from Honeypot as much information about attackers' activities as possible by stopping specific system calls (*syscalls*) on kernel level [13].

C. Advantages and disadvantages

All security technologies have some risk level. If knowledge and experience are power for attacker, they are also advantages for security professionals. By knowing the risk in Honeypots is possible to use knowledge to mitigate them and reduce disadvantages [2].

Advantages

- Small data sets – Honeypots can only monitor traffic that comes directly to them. They collect small amounts of data, but on the other hand may contain information of high value.
- Minimal resources – given that captures harmful activities, Honeypot requires minimum system resources – low-end system will be enough.
- Discovery new tools & tactics – Honeypots capture everything that gets start interaction with them.
- Encryption or IPv6 – Honeypot can operate also in encrypted or IPv6 environments or systems.

- Simplicity – Honeypots are very easy and flexible so they don't need complicated algorithms [4].

Disadvantages

- Risk of takeover – if attacker takes over Honeypot, he can exploit it to attack other systems inside or outside the system of deployment.
- Limited vision – Honeypots can only monitor traffic that comes directly to them.
- Discovery and fingerprinting – Honeypot has some expected characteristics or behavior so attacker using fingerprinting tool can identify it. Even simple error such as misspelled word in emulated service can act as Honeypot signature [4].

IV. IDS ARCHITECTURE USING HYBRID HONEYPOT

The main IDS weakness lies in ability to detect new type of attack, usage of different attack strategies or new tools, because they could not be detected. Proposed IDS is using hybrid Honeypot to reduce risk of detection failure, designing safety features of distributed systems and extensive data collecting. It minimizes any system intrusion threat, too. Hybrid Honeypot combines several tools: Snort, Dionaea and Sebek, which were describe earlier. Due to rapid response to attacks, proposed system analyzes all captured data in various formats and also warning reporting system.

Architecture consists of several clients and server. Clients collect information about attack and captured malware then send back to server. Server records and analyzes received data, issues warning and via web interface displays overall information. Architecture is designed to achieve effect of distributed information centralized management and to build complex distributed system of early warning.

A. Client architecture

Because of gathering data activities during attack are clients installed in the same domain. Depending on type of cyber-warfare activities are activated different system components for data sets collecting. Data sets are then sent to server for further analysis and subsequent updating of system security. Client architecture consists of three components (Fig. 4):

- Snort – monitors and filters packet during detecting

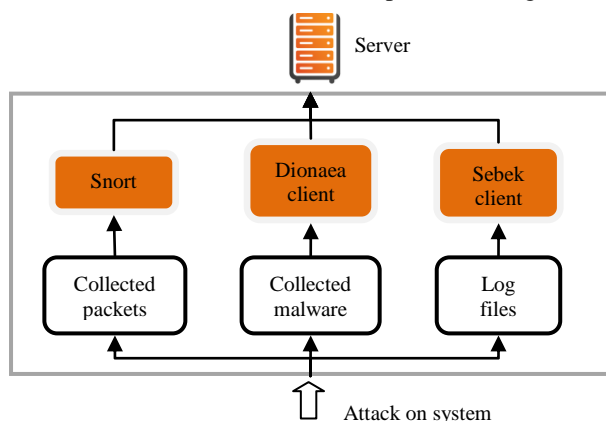


Fig. 4 Client architecture

intrusion. Identifies patterns & characteristics of attack, information and warning messages.

- Dionaea client – simulates general services and vulnerabilities, which attract attackers. Captures malware patterns & characters.

- Sebek client – denotes attacker behavior during interaction with Honeypot to log files.

B. Server architecture

Owing to centralization of collected data at the same time server is connected to multiple clients and is set to receive all outgoing messages, which are then stored in database (Fig. 3). Using individual reports interconnection it indicates attacker intention targeted to extensive computer or scanning attacks. The architecture of the Server consists of three parts, which outputs are normalized before they are going to be stored into database:

- Dionaea server – receives malware patterns sent by Dionaea client component.
- Sebek server – simultaneously receives and filters multiple data sources representing instruction or cohesion of sent data to store.
- Verification – modular design open-source hybrid system for detect intrusion using standard communication format. Can receive data from many clients and integrates disparate data formats.

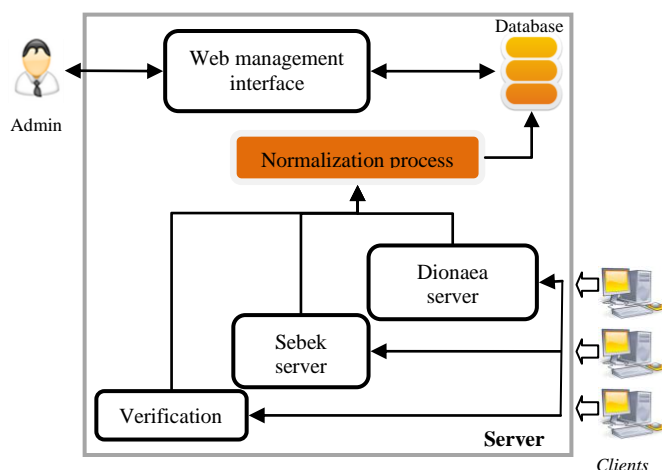


Fig. 3 Server architecture

Web server interface displays all attack analysis obtained from database. At the same time it monitors attack patterns and occurrences of unusual circumstances. In the event of their occurrence the concrete messages are highlighted to properly and in time response.

V. CONCLUSION

Security of information technologies is very essential in society that depends on information. Therefore in systems developing process is placed considerable emphasis on data and information sources protection. Protection of access, availability and data integrity represents basic safety features required for information resources. Any disruption of these properties would penetrate into the system and increase security risk. One way of defense is system that detects unusual and suspicious behavior – IDS. IDS major risk is represented by undetected penetration problem.

Advanced technology called Honeypot has huge potential for security community and it can achieve several objectives of other technologies, which making it almost universal. Usage of Honeypots represents a cost-effective solution to betterment organization security status.

Like any new technology Honeypot also have some

shortcomings that need to be overcome and removed. Neither of Honeypot many advantages it is not a panacea for breaching system security. Because it is used for gathering information about attacker and other threats is useful as an IDS detection mechanism.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0008-10.

REFERENCES

- [1] J. McHugh, A. Christie, J. Allen: "Defending Yourself: The Role of Intrusion Detection System," IEEE Software, IEEE Computer Society, October 2000, pp. 42-51.
- [2] L. Spitzner: "The Value of Honeypots, Part One: Definitions and Values of Honeypots," Security Focus, 2001.
- [3] L. Spitzner: "Honeypots: Tracking Hackers," Boston, USA: Addison-Wesley, Pearson Education, 2003, ISBN 0-321-10895-7.
- [4] S. Karthik, B. Samudrala, A. T. Yang: "Design of Network Security Projects Using Honeypots," Journal of Computing Sciences in Colleges, 2004.
- [5] R. Baumann, C. Plattner: "White Paper: Honeypots," Swiss Federal Institute of Technology, Zurich, 2002.
- [6] L. Vokorokos, A. Baláž: "Architecture of Computer Intrusion Detection Based on Partially Ordered Events," Vukovar, IN-TECH, 2010, pp 13-28, ISBN 978-953-307-047-6.
- [7] L. Vokorokos, A. Baláž: "Host-based Intrusion Detection System," IEEE 14th International Conference on Intelligent Engineering Systems, Budapest, 2010, pp. 43-47, ISBN 978-1-4244-7651-0.
- [8] L. Vokorokos, N. Ádám, A. Baláž, J. Perháč: "High-performance Intrusion Detection System for Security Threats Identification in Computer Network," Computer Science and Technology Research Survey, Košice, 2009, pp. 54-61, ISBN 978-80-8086-131-5.
- [9] L. Vokorokos, N. Ádám, A. Baláž: "Application of Intrusion Detection Systems in Distributed Computer Systems and Dynamic Networks," Computer Science and Technology Research Survey, Košice, 2008, pp. 19-24, ISBN 978-80-8086-100-1.
- [10] L. Vokorokos, A. Baláž, M. Chovanec: "Distributed Detection System of Security Intrusion Based on Partially Ordered Events and Patterns," Studies in Computational Intelligence 243, Budapest, September 2009, pp 389-404, ISBN 978-3-642-03736-8.
- [11] Snort [online]. Available on: <<http://www.snort.org>>.
- [12] Dionaea catches bug [online]. Available on: <<http://dionaea.carnivore.it/>>.
- [13] Sebek [online]. Available on: <<http://www.honeynet.org/tools/sebek/>>.

Section:
Electrical & Electronics Engineering
Oral form

Grid-tied Multilevel Inverter for Photovoltaic System

¹Marek PÁSTOR (2nd year)
Supervisor: ²Jaroslav DUDRIK

^{1,2}Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

¹marek.pastor@tuke.sk, ²jaroslav.dudrik@tuke.sk

Abstract—The paper presents a grid-tied cascade H-bridge inverter with predictive current control and reactive power control. The proposed 15-level cascade inverter consists of three H-bridge inverters with separated DC sources. At the output of the cascade inverter an L filter is used. The cascade inverter is controlled by the mean of RT-Lab. The predictive current regulator and one-phase synchronous reference frame phase-locked loop (PLL) are designed with help of Rapid Control Prototyping. The proposed control method uses a discrete model of the load to predict the behaviour of the system for each of 15 voltage levels of inverter's output voltage. Verification on a laboratory model is described.

Keywords—cascade inverter, phase locked loop, predictive current control, RT-Lab.

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. Whether it is a stand-alone photovoltaic (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. There are several types of PV inverters according to the topology. However, there is little experimentation with alternative inverter topologies [1]. The most widely used topology employs full-bridge (H-bridge) voltage source inverter.

The cascade H-bridge inverter is an alternative to the single H-bridge inverter in photovoltaic systems. Its advantages over the single H-bridge inverter are lower total harmonic distortion of the grid current (THDi) and THDu of the output voltage, requirements of smaller filters, ability to transfer more power and 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 [2]. Lower du/dt stresses of components in multilevel H-bridge inverter can help to meet these needs.

The lifetime of PV generators is in the range of 25 years and their reliability is high. However, the lifetime of typical inverter is in the range of 5 – 10 years (in 2006) [1]. It means that the inverter needs to be replaced several times during the lifetime of the PV generator. According to several biggest PV inverter producers, the PV inverter lifetime of 20 years cannot be achieved (mainly due to poor reliability of capacitors) and

the price of the inverter is more important than its lifetime. On the other hand, the cost reduction and the reliability increasing can be achieved by using new topologies of PV inverters [1]. The topologies of utility scale PV inverters are moving towards multilevel structures mainly because of lighter filtering components and better harmonic spectra [9].

Progress in the field of photovoltaics and increasing penetration of grid-tied distributed renewable energy sources leads to potential problems with those power sources into the grid. Small grid connected PV systems are usually connected to the low voltage grid. In Germany, which has more than 20 GW of grid connected PV systems installed, is approx. 80% of installed power fed to the low voltage grid. [8]

Such a large power supplied from renewable energy sources can significantly influence the low voltage grid. The PV plants are thus moving from pure grid feeders to sources at least partially responsible for power quality. As will be shown, the reactive power control can significantly influence the performance of grid-tied PV inverter.

II. CASCADE H-BRIDGE INVERTER

A. Multilevel Converter Topologies

Multilevel inverters have been used for many years in high-voltage, high-power applications. Their capability to divide the net voltage and power between several smaller cells and to produce higher quality voltage and current were the reason for their spreading in these areas. The most widely used topologies in industry are cascade inverter, diode-clamped neutral point clamped (NPC) inverter and capacitor-clamped (flying capacitor) NPC inverter.

Multilevel inverters usually need several separated dc sources which is one of the biggest problems they have. However, in the area of photovoltaics, the separated dc sources with galvanic isolation are not a problem. Even though, not all above-mentioned multilevel topologies are suitable for PV inverter. The diode-clamped NPC inverter has a complicated active power control and the capacitor-clamped NPC inverter has low efficiency when it has to transfer the active power [3].

B. Cascade H-bridge Inverter

A single-phase structure of a general 7-level cascade H-bridge inverter is shown in Fig.1. The number of output phase voltage levels n is defined by:

$$n = 2d + 1 \quad (1)$$

where:

d – is the number of separated dc sources.

However, by using asymmetrical DC voltages at the cascade inverter input, where the next DC source voltage level is two-times the previous DC voltage level, the number of output voltage levels can be increased to:

$$n = 2^{d+1} - 1 \quad (2)$$

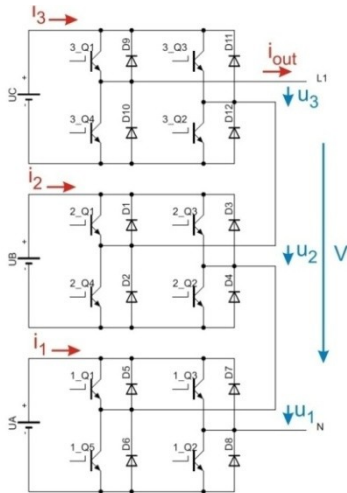


Fig. 1. Single-phase cascade H-bridge inverter with three separated DC sources ($U_A = 40V$, $U_B = 20V$, and $U_C = 10V$), capable of creating 15 voltage levels at its output.

Each H-bridge inverter can create positive, negative or zero voltage on its output with magnitude equal to the dc source voltage. Thus there are 15 possible combinations for the cascade H-bridge inverter with 3 separated dc sources (2).

III. DISTRIBUTED GENERATION SYSTEM

Grid connected PV system can influence the power flow in the point of coupling in several ways.

In Fig. 2 each customer is drawing certain apparent power from the grid. The customer 2 has installed PV plant. If the PV plant is supplying only real power which is high enough when compared to the power S_2 drawn from the grid, then customer 2 is drawing the power from the grid with poorer power factor when compared to the scenario without the PV plant. The reason is that a portion of the real power of S_2 is supplied by the PV plant but the apparent power drawn from the grid stays the same. In the case that the PV plant is capable to cover the whole active power of S_2 only the reactive power would be drawn from the grid.

On the other hand, if the PV plant output power is higher than the consumption S_2 , the power supplied from the PV plant would cause the increase in voltage in the point of coupling.

The solution to these problems is reactive power control in the PV inverter. In the first case, the PV inverter capable of reactive power control would improve the power factor at which is the power drawn from the grid. And in the second case, the consumption of reactive power by PV inverter would decrease the voltage in the point of coupling.

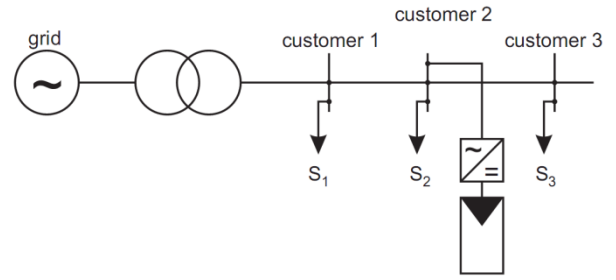


Fig. 2. PV plant as distributed power source

When considering the control structure for grid connected PV systems, there are two possibilities to choose from. Either voltage or current control can be chosen. The basic difference between them is that voltage controlled PV systems is capable of operating without the grid whilst the current controlled PV system is not and it requires strong grid as reference. With the voltage control there is a connection between active and reactive power delivered to the grid and the reactive power is not fully independent from the real power. On the other hand, with the current control, there is a fully independent control of reactive power.

The current regulation structure of a grid connected PV system is depicted in Fig. 3. One phase system is considered. The inputs to the regulation structure are two currents in synchronous reference frame: i_{dref} and i_{qref} . The synchronous reference frame is synchronized to the grid by mean of one-phase synchronous reference frame PLL. The PV system is current controlled. The output from the reference current generator is fed to the predictive current controller which is together with the coder responsible for generating gating pulses for the 15-level cascade inverter. The cascade inverter is connected to the grid through the grid filter which is for now just a simple L filter.

The i_{dref} current is responsible for active power transfer from the PV system to the grid whilst the i_{qref} is responsible for reactive power flow between the PV system and the grid. Thus by changing the i_{dref} current (by maximum power point tracking algorithm), the power extracted from the PV generator can be changed.

IV. CASCADE INVERTER CONTROLLER

The cascade inverter controller (Fig. 3) consists of several parts: grid measurements, PLL, reference current generator, predictive current controller and coder.

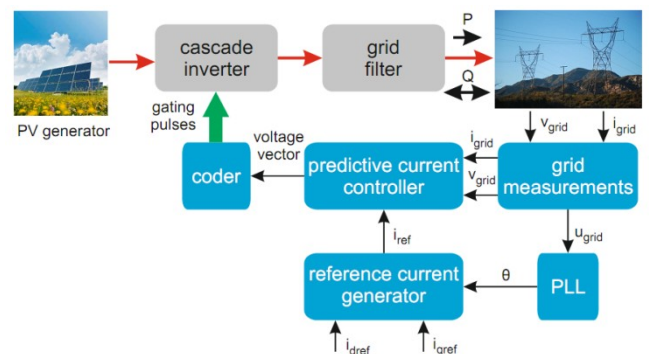


Fig. 3 Current control of grid connected PV system

A. Phase Locked Loop

The PLL (Phase Locked Loop) is a mean how to synchronize the grid connected system to the grid voltage. The synchronization is needed for power factor control. PLL synchronization techniques can be divided into two groups: open loop and closed loop. The most widely used technique in three-phase system is synchronous reference frame PLL (SF-PLL). The SF-PLL has good performance with the grid which is not highly distorted [10].

The SF-PLL is based on direct Clark transformation of the three-phase system into two-phase system and the subsequent Park transformation into synchronous reference frame. Thus two voltages v_d and v_q are produced. One of these voltages is by a mean of PI controller set to zero which results in the reference being locked to the grid. The output from the PLL is a phase angle which is used to generate the reference three-phase currents through direct Park transformation ($dq \rightarrow \alpha\beta$) and subsequent reverse Clark transformation ($\alpha\beta \rightarrow abc$).

The SF-PLL can be used for one-phase systems as well. However, it is not possible to use the direct Clark transformation because only one voltage is presented. The solution is to create artificial two-phase system based on the one-phase grid voltage [10][11].

The property of the stationary reference frame is that two voltages v_α and v_β are orthogonal. If the grid voltage corresponds to the v_β voltage, then the v_α can be created as follows:

$$\begin{bmatrix} v_\alpha \\ v_\beta \end{bmatrix} = \begin{bmatrix} v_{\text{grid}} (\omega t - \pi/2) \\ v_{\text{grid}} (\omega t) \end{bmatrix} = \begin{bmatrix} V_m \sin(\omega t - \pi/2) \\ V_m \sin(\omega t) \end{bmatrix} \cong \begin{bmatrix} -V_m \cos(\omega t) \\ V_m \sin(\omega t) \end{bmatrix} \quad (3)$$

There are several possibilities how to create the 90 degrees phase shift of the grid voltage to produce the v_α voltage (e.g. storage elements, filters). One of them is to use second-order low-pass filter [10][11]. Based on the comparison in [10] the best performance is achieved by PLL with second order filter artificial voltage generator and SF-PLL. When the input voltage v_{grid} passes through the second-order low-pass filter, where the damping ratio $\zeta = 1/\sqrt{2}$, the undamped natural frequency ω_n has the same value as the estimated frequency, a signal with a phase-angle difference of $\pi/2$ and amplitude of $V_m/\sqrt{2}$ is obtained [10]:

$$u_\alpha = -\sqrt{2} \frac{V_m}{\sqrt{2}} \sin\left(\omega t - \frac{\pi}{2}\right) = V_m \cos(\omega t) \quad (4)$$

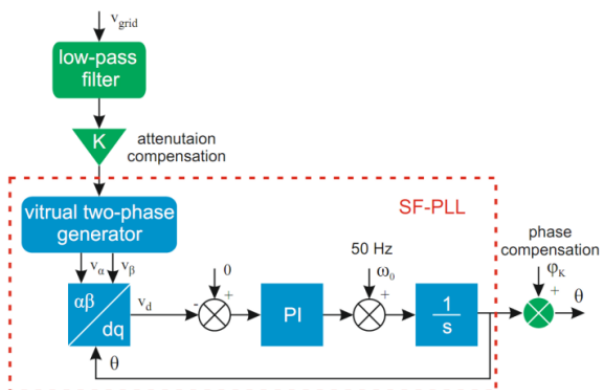


Fig. 4 One-phase synchronous reference frame PLL (virtual two-phase generator is 2nd order low-pass filter)

The one-phase synchronous reference frame PLL is shown in Fig. 5. The input is a grid voltage which passes through a low-pass filter. The amplitude and phase of that filter is compensated, as suggested in [10].

The PLL is implemented as discrete model and filters' coefficients are obtained by Tustin transformation.

B. Reference Current Generator

The reference current generator consists simply of direct Park transformation. The i_α is set as reference current i_{ref} for current regulator.

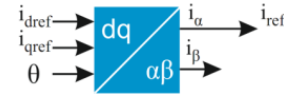


Fig. 5 Reference current generator

C. Predictive Current Controller

There are various control techniques to control the output of the inverter. When using the PWM control there is a need to linearise the model of the inverter and this control technique can lead to cascade regulation structure which has slow response time. The predictive control offers the possibility to control the inverter's output current and voltage with high dynamics without the need to face the problem of non-linear nature of semiconductor power converters [6].

In [7], authors use trajectory based predictive control for current control of three-level diode-clamped NPC inverter. This control technique has been adapted to the proposed 15-level cascade H-bridge inverter. This control technique can be broadly classified as trajectory based.

The basic principle of used predictive control technique is that the cascade H-bridge inverter can create only limited number of voltage levels at its output.

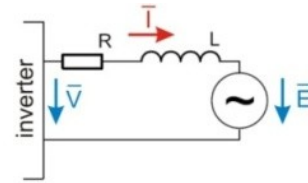


Fig. 6. The L filter between the inverter's output and the grid used to decouple the output voltage and the grid and to filter higher harmonics.

In Fig. 6 is depicted a PV inverter connected to the grid through simple L filter. The variable of interest is the current supplied to the grid I . This current is influenced by the inverter voltage V . The goal is to predict the behaviour of the load current I for each possible voltage vector generated by the inverter cascade. For the system in Fig. 6, the future value of the load current can be predicted from the model of the system as:

$$i(k+1) = \frac{T_s}{L} (v(k) - e(k)) + i(k) \left(1 - \frac{RT_s}{L}\right) \quad (8)$$

The (8) is used to predict the future value of the load current. For the trajectory based predictive control there is a need to create the trajectory which will the controlled variable follow. However, the future value of the reference current

$i^*(k+1)$ is unknown. In order to determine the next value of the reference current, in [7] Lagrange quadratic extrapolation is used, that at each point assume the corresponding value of the function:

$$i^*(k+1) = 3i^*(k) - 3i^*(k-1) + i^*(k-2) \quad (9)$$

For predictive control, there is a need to create the cost function which will be evaluated in each sampling time and will define the behaviour of the system. The function can be chosen as a filter to remove certain harmonics and so on. [6] The cost function was chosen as:

$$z(k) = |i^*(k+1) - i(k+1)| \quad (10)$$

The controller structure is shown in Fig. 7. The inputs are grid voltage and current and the reference current from the reference current regulator. The discrete model of the system is calculated for each voltage vector and the voltage vector that minimise the cost function is chosen. The output of the predictive current regulator is the desired voltage vector which is fed to the coder. The coder is responsible for the control of switching states of H-bridges to create the desired voltage level at the inverter's output.

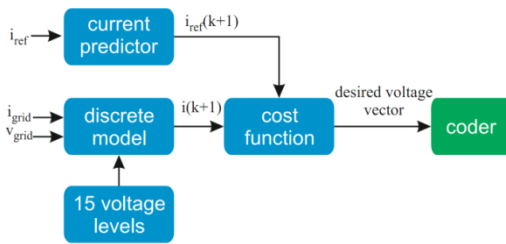


Fig. 7 The predictive current controller

V. EXPERIMENTAL RESULTS

With measurements on the 15-level cascade H-bridge inverter, the tracking capability of the load current and the power factor control were verified. The full power (3.6 kW, ~230 V/ 50 Hz) laboratory model is built. For measurements, the system parameters were set to: $R = 5 \Omega$, $L = 7 \text{ mH}$, $T_s = 100 \mu\text{s}$, $U_A = 40 \text{ V}$, $U_B = 20 \text{ V}$, $U_C = 10 \text{ V}$ (voltage sources) as just designed control technique was verified. The inverter was connected to the single-phase grid ~35V/50Hz and controlled by RT-Lab.

The reference current was set as $i_{\text{dref}} = 2 \text{ A}$, $i_{\text{qref}} = 0 \text{ A}$, thus the power factor should be 1. The active and reactive power as well as phase shift was measured by power analyzer. The real phase shift was 2 degrees into inductive region thus the reactive power was consumed by the inverter (Fig. 8).

VI. CONCLUSION

In this paper the 15-level grid-tied cascade H-bridge inverter with predictive current control technique is presented.

From the experimental results can be concluded that the proposed current regulator and PLL synchronisation work properly.

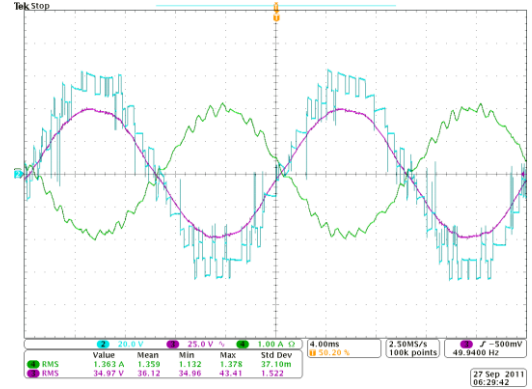


Fig. 8 Output current of cascade inverter with current control: CH2: inverter voltage (inverted), CH3: grid voltage, CH4: grid current ($T = 100 \mu\text{s}$, $R = 5 \Omega$, $L = 7 \text{ mH}$, $U_A = 40 \text{ V}$, $U_B = 20 \text{ V}$, $U_C = 10 \text{ V}$), $\text{THDi} = 4\%$, $P = 50 \text{ W}$, $Q = -1,3 \text{ W}$, $\phi = 2^\circ$ inductive, $i_{\text{dref}} = 2 \text{ A}$, $i_{\text{qref}} = 0 \text{ A}$

Also the reactive power control is ready to be used with full output power of the inverter. Connection to the PV generator and design of MPPT (Maximum Power Point Tracking) control, which is essential for a PV inverter, is needed to be designed. Also the problem with unloaded dc sources (partial PV generators in real system) needs to be solved. Also replacing the simple L filter with a filter of higher order (LCL filter) is one of the next challenges. The used cost function is simple and by creating more sophisticated cost function, the control technique can be improved. However, the work shows promising results and by using computer control and RT-Lab, the controller can be easily modified.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0185-10.

REFERENCES

- [1] NREL: A Review of PV Inverter Technology Cost and Performance Projections, [Online].
- [2] "Current Demand of High Performance Inverters for Renewable Energy Systems" [Online].
- [3] S. Khomfoi, L. M. Tolbert: Multilevel Power Converters, kapitola from H.Rashid: Power Electronics Handbook, second release, Elsevier, 2007, ISBN10: 0-12-088479-8, pp. 451 – 482.
- [4] M. Calais, V. G. Agelidis and M Meinhardt, "Multilevel Converters for Single-Phase Grid Connected Photovoltaic Systems: An Overview, [Online].
- [5] 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].
- [6] R. Kennel, "Predictive Control of Inverter Supplied Electrical Drives", in Power Electronics Specialists Conference, 2000. PESC 00. 2000 IEEE 31st Annual, 2000, pp. 761 - 766 vol.2.
- [7] G. S. Perantzakis, F. H. Xepaps, S. A. Papathanassiou, S. N. Manias, "A Predictive Current Control Technique for Three-Level NPC Voltage Source Inverter", in Power Electronics Specialists Conference, 2005. PESC '05. IEEE 36th, 2005, pp. 1241 – 1246.
- [8] SMA, "Technology Compendium 3.3 – PV Grid Integration", [Online], 2011
- [9] Yaosuo Xue, K.C. Divya, G. Gripenstrog, M. Liviu, S. Suresh, M. Manjrekar, "Towards Next Generation Photovoltaic Inverters" [Online]. IEEE, 2011
- [10] J.-W. Choi, Y.-K. Kim, H.-G. Kim, "Digital PLL control for single-phase photovoltaic system" [Online]. IEEE, 2006
- [11] B. Meersman, J. De Kooning, T. Vandoorn, L. Degroote, B. Renders, L. Vandeveld, "Overview of PLL methods for distributed generation units" [Online]. IEEE, 2010

Design of the 5-bit Flash A/D Converter for UWB Applications

¹Martin LIPTAJ (3rd year PhD.), ²Tomáš HRÁŠOK (2nd year Ing.)

Supervisor: ³Pavol GALAJDA

^{1,2,3}Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

¹martin.liptaj@tuke.sk, ²tomas.hrasok@student.tuke.sk, ³pavol.galajda@tuke.sk

Abstract—The design of the 5-bit flash analog to digital converter for UWB applications that supports sampling rates up to 50 Msample/s is presented. Output codes of these converter are gray and binary. Input voltage range is designed from approximately -900mV to 0V and input frequency bandwidth up to 300MHz. The A/D converter will be a part of an ASIC circuits family designed for implementation in the recent UWB sensor system architecture.

Keywords—A/D converter, 5-bit, flash, SiGe BiCMOS, UWB

I. INTRODUCTION

Current trend in communications is to place all system to one chip (System on Chip - SoC) or all system to one package (System on Package - SoP). It is well known, that this monolithic integration brings a broad scale of benefits, e.g. typically in the overall system performance, reliability improvements due to reduced interconnect and package parasitics, smaller package count, reduced supply power requirements because fewer high-frequency signals that require usually 50Ω interfaces are routed off chip and so on. On other side the SoC and SoP architectures comprises broadband high frequency relatively high power outputs and sensitive input devices on one substrate or package respectively. Thus, the undesired signal coupling can degrade the performance of the sensitive receive circuitry and so of whole system.

Up till now gained knowledge on the field of monolithic RF-device integration in combination with the push for lower cost, smaller device size and more features has motivated the effort towards increasing of sensor ASIC complexity (e.g. towards single ASIC for MIMO UWB sensor array application). So the first experimental design combining all key M-Sequence sensor system high-frequency analog subcomponents on one silicon substrate was undertaken [1]. For the following signal processing it is necessary to convert analog data from sensor system to digital data. This conversion is provided by Analog to Digital (A/D) converter. For minimize of interference, conversion loss etc. it is useful connect A/D converter and sensor system together in one package. For this purpose was developed and designed prototype of 5-bit flash A/D converter.

The presented paper consists of a three parts. In the first part, requirements and reasons for the choice of architecture are presented. The second part deal with the presented A/D converter design, especially with its parts as a comparator, delay line, input and output buffers and others. Finally, a several simulations of A/D converter and preliminary measurements are described.

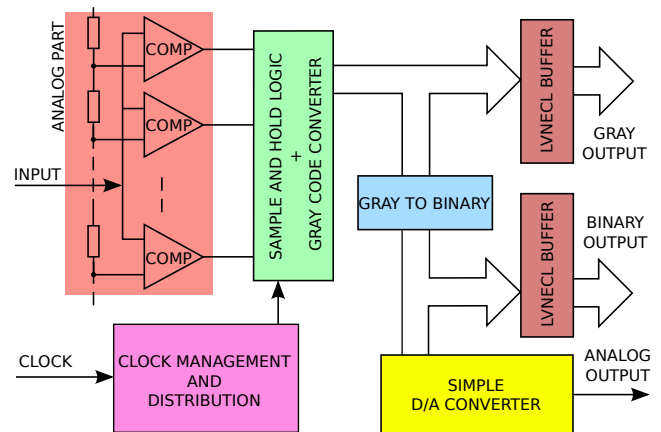


Fig. 1. Block diagram of proposed A/D converter

II. THE DESIGN OF A/D CONVERTER

A. Choosing the structure of the converter

In the UWB devices it to capture signals from the communications network with the frequency of the order of hundreds of MHz is needed. For these requirements the parallel converters that provide very fast digitizing of analog signal are the best solution. These converters are also called as "flash" converters. They can capture the input signal and generate all the binary output practically at the same time [2]. Their function is based on comparing the level of the input signal with the reference voltage from resistor ladder. If you increase the resolution of the converter, exponentially growing number of resistors in the resistor ladder and the number of comparators and logic gates. This implies the biggest disadvantage of the parallel converter, which is complex design layout. For the resolution of the N-bit converter is needed exactly 2^{N-1} comparators [3]. Despite of this disadvantage of the parallel converters are often used, especially in areas of high-frequency band.

For the UWB radar developed by the TU Ilmenau the design of 5-bit parallel converter has been chosen. This design is a compromise that takes into account the technology of chip manufacturing, power consumption, design complexity of IC and the necessary resolution of the A/D converter.

B. Technology

Generally there are several technological possibilities for monolithic integration of the ASICs for high frequency system architectures. Industrial very preferred are silicon based

processes, where the silicon can be cheaply grown on the very large defect free crystals yielding many low cost ICs per wafer. Silicon (Si) is a relatively good thermal conductor allowing efficient removal of power dissipated as heat. GaAs and InP on the other hand are commonly grown in smaller wafers with higher defect densities and exhibit lower thermal conduction. In digital and also in certain level of analogue circuits, the use of a large number of both n-type and p-type devices are required. Si based processes offers the availability of PNP and NPN transistors (Si can be easily doped to make both n-type and p-type devices). Additionally, the silicon technologies are available for broader scale for customers, what is not omissible fact. So silicon based processes seems to be better suited for highly integrated circuits.

A derivative of Si CMOS, so called SiGe BiCMOS process, with its bandgap engineering has made it possible to achieve low cost and relatively high volume production (the property of Si CMOS) for high frequency applications. SiGe HBTs have a lower and graded bandgap, accelerating majority carriers across the base. This improves high frequency performance with higher maximum unity current gain cut-off frequency (f_t) and maximum oscillation frequency (f_{max}) in region beyond 300GHz to meet analog requirements [3][4].

CMOS devices on the other hand would be developed and scaled mainly for digital logic by thinning the gate oxide, shortening the channel length, lowering the supply voltage and hence reducing power consumption. This simplifies the CMOS process side and reduces its complexity and cost. According to the all afore mentioned reasons the SiGe BiCMOS technology is favoured for implementation of ASIC family for recent A/D converter for UWB sensor system architectures, because in the future development the high-frequency analog and complex digital circuits may be integrated on the same chip [4].

One of the promising technologies for this undertaken is 0.35 μ m SiGe-BiCMOS process from Austriamicrosystems (AMS), Austria [5]. This technology is based on the proven 0.35 μ m mixed-signal CMOS process and includes an additional high performance analog oriented SiGe HBT transistor module. It supplies wide range of high performance RF applications up to 20Gb/s. Bipolar module cut-off transient frequency f_t is more than 60GHz and maximal frequency f_{max} is more than 70GHz [5]. Based on the selected SiGe BiCMOS technology, the negative supply voltage -3.3V for A/D converter has been chosen.

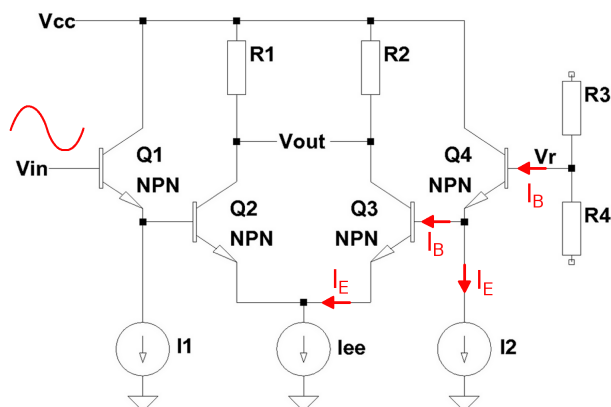


Fig. 2. DC bowing of comparator

C. Building blocks

A conventional A/D converter consist of three main parts: Sampling circuit, quantizer and decoder [3]. In the flash A/D converters are these blocks realized by 2^{N-1} comparators, resistor ladder and converter of the thermometer to some digital code (Fig. 1.). In the design of parallel A/D converter is given most importance for sampling circuits and comparators. The role of the sampling circuit is to capture the current value of the input signal and keep this value for time necessary for the quantization process. In our converter design, we missed this circuit, as it will be implemented externally, outside of the chip. The reference voltage is realized by on-chip resistor ladder. This resistor ladder is created from 30 series-connected resistors and powered by the source of voltage 860mV. Voltage drop of 29mV occurs on each resistor. It is clear, that each comparator must be capable to invert his output at the difference between the input and reference voltage less than 29mV. The comparator was realized as differential driver based on a bipolar transistors, which are powered by current source, because of a linear characteristic. Current source is made by a current mirror composed also by bipolar transistors. Commonly stage provides a high degree of differential sensitivity and low capacity of input pins. Disadvantages consist in kickback noise, AC and DC deviation [3].

For control of the input differential stage of comparator is needed considerable current flowing into the comparator input. This current brings deviation into reference resistor ladder. It is also called the DC bowing. This deviation can be corrected by correct design of resistor ladder. Principle of DC deviation is shown in Figure 2.

When using parallel A/D converter in the high frequency range of the hundreds of megahertz, non linear AC bowing occurs as well. This occurs in parasitic capacity formed by the internal structure of transistors and capacity of wires. Figure 3. shows the parasitic capacity of the comparator internal structure formed by bipolar transistors. Even though bipolar transistors have a relatively low input capacity, it is clear that increasing the accuracy of the converter (increasing the number of comparators) increases directly the input parasitic capacity between the comparator inputs V_{in} and V_r . This leads to significant deviations of the reference voltage on resistor ladder. One possible solution is to create "stronger" resistor ladder that brings a higher current consumption of the converter. AC deviation can be partially corrected also with capacitors connected on each of reference voltages [6].

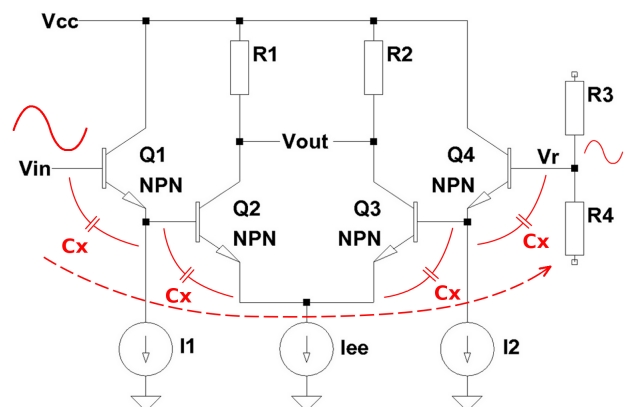


Fig. 3. AC bowing of comparator

The output of each comparator is connected to D flip-flop, that operates together with clock signal as memory cell (Fig. 1.). This element provides synchronization for output signals of comparators. As a primary transmitter code was selected Gray code, which produces less errors in case of failure in the output thermometer code of comparators. Next advantage is a simpler logical structure [3]. At output of Gray code converter the binary code converter is additionally connected. It is consist of four XOR gates.

A simple D/A converter internally connected to the binary outputs is placed on the proposed A/D converter chip. D/A converter is based on R/2R resistor network topology on which output is connected a simple smoothing filter. The whole structure of A/D converter is operated at negative CMOS voltage level of -3.3V. Outputs of binary code and Gray code are equipped with voltage converters that adapt standard LVNECL logic levels.

The layout of the proposed A/D converter chip shown in figure 4. The chip dimension selected for implementation of ASIC was 1950x950 μm .

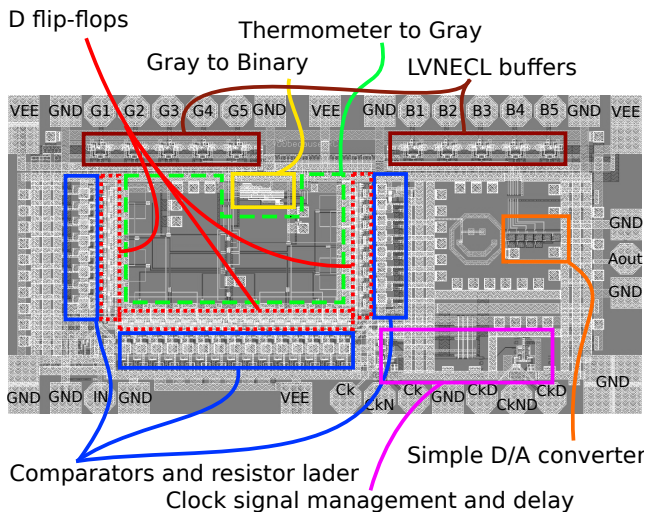


Fig. 4. The layout of A/D converter chip

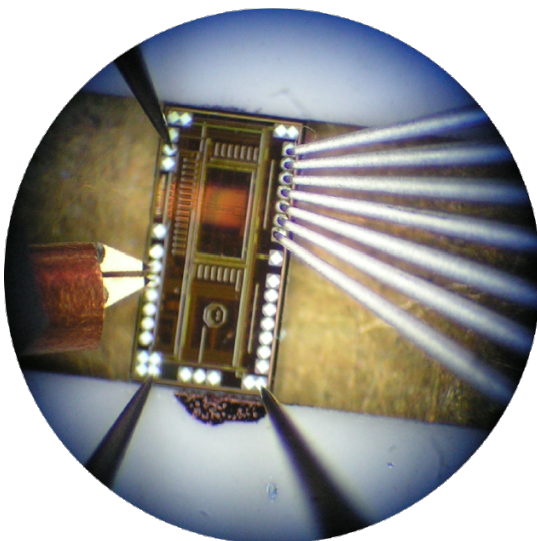


Fig. 5. The measurement on wafer probe station

III. THE MEASUREMENT OF THE A/D CONVERTER

The size of the contacts pads of the chip is 100 μm . The distance between the pads is less than 10 μm . Such a chip size can be tested only using a special high-frequency probes for measurement and power. This probes meet the standard impedance parameters and the standard distances.

A first step of measurement was the check of power consumption. The chip has four pads for connecting of power supply and thirteen pads for connecting on the ground. However, due to lack of space around the chip we connected only one power and one grounding pad. The chip has 108mA of current consumption at -3.3V of supply voltage. The designed power consumption of whole chip was 600mW, which means 182mA. The difference was caused mainly by loss voltage on DC supply probes. Clock frequency of 100kHz signal was connected to the chip by 50 Ω RF probe Cascade ACP40-SG-100 μm (Fig. 5., on the left). Digital Gray code output was captured by multi-probe Cascade DCQ07, that has seven pins (Fig. 5., on the right).

Because of not enough space around the chip, we connected the input signal through probe designed for DC power supply. The input signal was sine wave with amplitude of 600mV and a frequency of 1kHz. Expected output logic levels of digital signal of A/D converter were -1.1V and -1.75V. After connecting all the power and signal pins, we began to increase the supply voltage. Chip started to generate an output signal at -3.86V supply voltage and 132mA of supply current. The binary output signals were connected to the oscilloscope Lecroy MSO-104MX in order to check the decision voltage levels for logic analyser. Then we connected all five outputs to logic analyzer, where we set the output level. In Fig. 6. and Fig. 7. the measured results can be seen. The yellow lines represents the sine wave input signal, the red lines represent the output gray code and binary code respectively. The upper red line represents the MSB, the lower red line represents the MSB. As can be seen on these figures, there are a glitch of digital signal, which are probably caused by improperly engaged power supply for A/D converter. As a result of insufficient powered chip by one DC probe was created voltage differences. These differences affect the functionality of comparators. There were also measured different offset voltages of the digital outputs of the converter. Unbalanced input signal also introduces considerable noise and interferences into the measurement. This signal was not adapted by optimal impedance. Measurements of A/D converter was realized at several frequencies. The converter output generated relevant data and clock signals also at 10MHz. Maximum input and clock speed has not yet been measured.

Fig. 8. shows the output signal from simple D/A converter with input frequency of 10kHz. During the measurement the converter was placed on a large metal surface (Fig. 5.). Chip temperature was measured by a non-contact laser thermometer and did never exceed 40°C.

The proposed A/D converter will pass additional measurements. It is necessary to add sense lines to ensure a stable operating voltage on-chip. The chip will be also bond on to the PCB. After that the additional measurements will be performed.

IV. CONCLUSION

The design and preliminary measurement results of the prototype of 5-bit flash A/D converter design, which is a

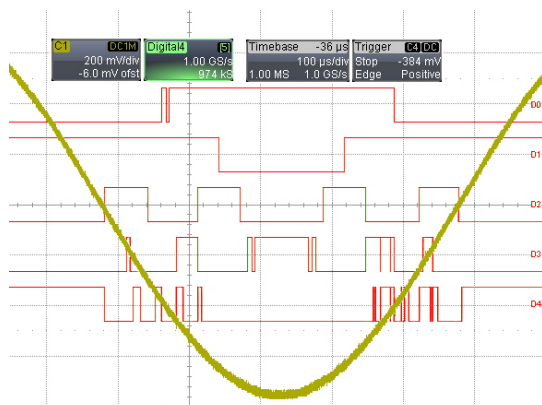


Fig. 6. The measured signal from gray code output

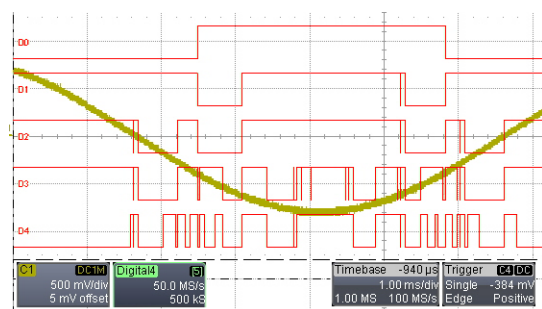


Fig. 7. The measured signal from binary code output

part of ASIC kit for UWB sensor system are presented. With the help of modern CAD software was designed and manufactured an experimental A/D converter, which is still in phase of measurement. We have gained valuable knowledge about design and measurement of converters.

ACKNOWLEDGMENT

The authors will acknowledge Martin Kmec and also whole team at EMT, TU Ilmenau for support. This work is also the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] M. Kmec, M. Helbig, R. Herrmann, P. Rauschenbach, J. Sachs, and K. Schilling, "M-sequence based single chip uwb-radar sensor," *Ultra-Wideband Short-Pulse Electromagnetics 10, ANTEM/AMEREM 2010 Conference*, 2010, Ottawa, ON, Canada, July 2010.
- [2] H. Pan and A. A. Abidi, "Signal folding in a/d converters," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS: REGULAR PAPERS*, pp. 3–14, 2004, VOL. 51, NO. 1, JANUARY 2004.
- [3] B. Razavi, "Principles of data conversion system design," 2010, ISBN 0-7803-1093-4.
- [4] 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*, pp. 60–64, September 2010, Velké Losiny, Czech Republic. Ostrava, VŠB-TU, ISBN 978-80-248-2261-7.
- [5] Austriamicrosystems, "Global leader in the design and manufacture of high performance analog ics (integrated circuits)," *Available on internet 10.03.2012.*, <http://www.austriamicrosystems.com>.
- [6] H. Yu, "A 1v 2.5gs/s 8-bit self-calibrated flash adc in 90nm gp cmos," *University of California, Los Angeles*, 2008, ISBN 0549723951, 9780549723950.

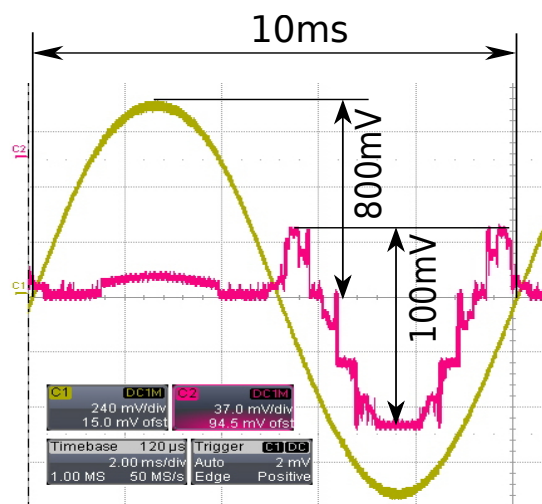


Fig. 8. The measured signal from simple D/A converter

Calculation of Inductance of Double-Layer Planar Spiral Coil Using COMSOL Multiphysics

¹Matúš OCILKA, ²Oleksii KRAVETZ, ³Oleksander SHUTKA
Supervisor: ⁴Dobroslav KOVÁČ

^{1,4}Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

^{2,3}Institute of Electromechanics, Energy Saving and Control Systems, Kremenichuk Mykhailo Ostrohradskyi National University, Ukraine

¹matus.ocilka@tuke.sk, ⁴dobroslav.kovac @tuke.sk

Abstract— the aim of this paper is to estimate the inductance of double-layer flat spiral (pancake) coil. First, the analytic approach of single-layer, then double-layer coil is presented and then the calculation is made using finite elements of solution. For this purpose the software package called COMSOL Multiphysics with AC-DC Module was used. Designed coil is created in 2D axisymmetric space dimension and results are compared with analytic solution.

Keywords—spiral coil, inductance, COMSOL Multiphysics

I. INTRODUCTION

For calculating inductance of spiral coils there are several analytic approaches and models. These approximations are usually quite complex and in some cases the analytic approach could be difficult and with increasing complexity of geometries of some coils, involving mutual inductance, surroundings and self-inductance of the conductor, these are sometimes difficult to estimate. Some formulas calculate the spiral inductors for RF-IC applications with inductance less than 100 nH with single layer geometry [3], [5]. For example in induction heating or RF-telecom designs applications which need the inductances more than μH or mH. Also accuracy in designing a spiral inductor is important because once it is built for example on printed circuit board it is difficult to modify the inductor [4]. Thus using of finite element method can be good alternative of solving such problems.

COMSOL Multiphysics is software package that solves systems of coupled three-dimensional partial differential equations. This software can be used to model different physical phenomena including electromagnetics. It uses the finite element method of solution and can model the complex problems, geometries or material properties in 2D or 3D representation for better understanding some phenomena. In this case COMSOL Multiphysics solves the magnetic fields surrounding the coil that is placed in electrical circuit and calculates the inductance of this coil.

II. PROBLEM FORMULATION

The spiral coil is a spiral – shaped formation made of copper. This copper spiral is electrically conductive showing an electrical resistance for the direct current. For the

alternating current the inductor shows, except the electrical resistance, also a property called inductance, which describes the capability of producing a magnetic field around the inductor. From the electrical point of view the inductance acts as complex impedance being proportional to the frequency of alternating current. The ideal inductor shows only the inductance while its electrical resistance is negligible [1].

III. THE COIL GEOMETRY

A. Single-layer spiral coil

The proposed single layer coil is shown in Fig.1. This inductor is physically described by the following parameters: d_{in} as the inner diameter [mm], d_{out} is outer diameter [mm], s expresses the spacing between turns [mm], w is the width of conductor [mm], and h is height of conductor [mm].

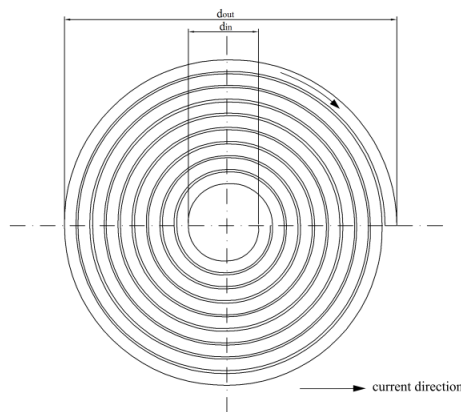


Fig. 1. Designed planar spiral coil

B. Double-layer coil

Double-layer inductor has the same geometry as a single layer coil, but it consists of two same coils with the distance of both spirals is $d = 1.5\text{mm}$. The geometry of coil is depicted in Fig.2.

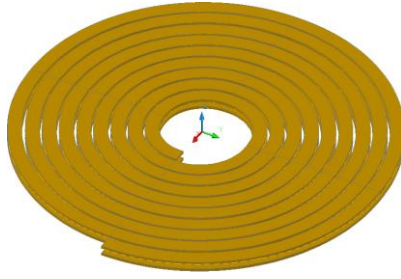


Fig. 2. Double-layer coil

IV. COIL INDUCTANCE

A. Inductance of single-layer coil

Calculation of inductance value of such coil can be done using following equation [4].

$$L = \frac{\mu_0 N^2 d_{avg} C_1}{2} \left(\left(\ln \frac{C_2}{\rho} \right) + C_3 \rho + C_4 \rho^2 \right) \quad (1)$$

where ρ is the fill ratio

$$\rho = \frac{d_{out} - d_{in}}{d_{out} + d_{in}} \quad (2)$$

and d_{avg} is average diameter

$$d_{avg} = \frac{d_{out} + d_{in}}{2} \quad (3)$$

Constants C_1 – C_4 are shape dependent constants and their values according to [4], are $C_1 = 1.00$, $C_2 = 2.46$, $C_3 = 0.00$, $C_4 = 0.20$, and $\mu_0 = 4\pi \times 10^{-7} \text{ H.m}^{-1}$.

B. Inductance of double-layer coil

The inductance of double-layer coil is given by following equation

$$L_{tot} = L_1 + L_2 \pm 2M \quad (4)$$

Where L_1 and L_2 are the self-inductances of both coils and M is the mutual inductance between coils. Mutual inductance is difficult to calculate because it depends of many factors as follows: distance between coils, geometry, and mutual geometry of particular conductors. This inductance increases or decreases total inductance of coil depending on currents directions passing through the coils. In our case the currents of both coils pass through coils in the same direction. The mutual inductance can be determined as:

$$M = K_c \sqrt{L_1 L_2} \quad (5)$$

Where K_c is the coupling factor and its value must lie in interval $<0;1>$. The coupling factor can be determined using following formula[4]:

$$K_c = \frac{N^2}{0.64 \cdot (Ax^3 + Bx^2 + Cx + D) \cdot (67N^2 - 5,84N + 65)} \quad (6)$$

Where $x=d$ is the mutual distance of both coils, constants A, B, C, D according to [4], are: $A = 0.184$, $B = -0.525$, $C = 1.038$, $D = 1.001$.

V. MODEL OF COIL

The model of coil is shown in Fig.3 and it is modeled in COMSOL Multiphysics in AC/DC module. The model is created in 2D axisymmetric space dimension and coil is surrounded with an air domain and its boundary serves as a magnetic insulation.

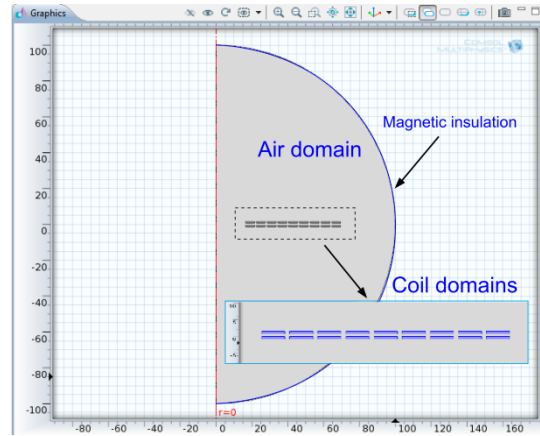


Fig. 3. Double-layer coil in COMSOL

In Fig. 4 is shown the 3D representation of such coil. This model is not exact spiral geometry and it was adapted for solving in 2D axisymmetric space dimension because spiral-shaped formation of this coil is not axisymmetric. One layer of coil is created from 9 rings as depicted in Fig. 4.

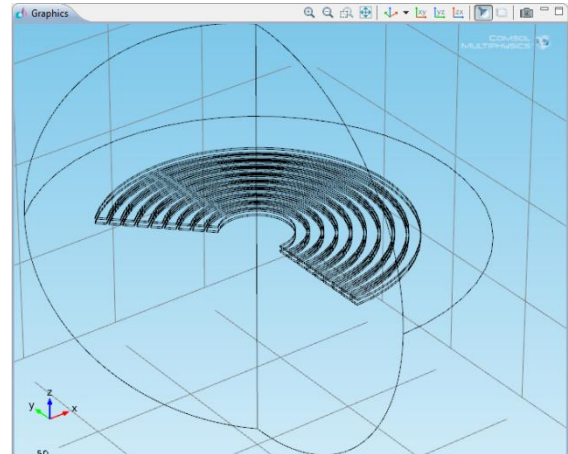


Fig. 3. Double-layer coil in 3D representation

VI. RESULTS

The parameters used in calculation and simulation are:

Number of turns:	$N = 9$
Outer diameter:	$d_{out} = 142 \text{ mm}$
Inner diameter:	$d_{in} = 30 \text{ mm}$
Width of conductor:	$w = 5 \text{ mm}$
Spacing between turns:	$s = 1 \text{ mm}$

Height of conductor: $h = 0.8\text{mm}$

Distance of coils: $d = 1.5\text{mm}$

A. Analytic calculation

The inductance of one coil using (1) is $L_1 = 6.19\mu\text{H}$. Both coils have the same geometry, so the inductance of second coil must be the same $L_2 = 6.19\mu\text{H}$. Using (4) we can calculate the total inductance of double-layer coil. The coupling factor according to (6) is $K_c = 0.855$ and mutual inductance (5) is $M = 5.29\mu\text{H}$. Then the total inductance using analytic approach is $L_{\text{tot}} = 22.90\mu\text{H}$.

B. Simulation results

For simulation the model shown in Fig.3 was used. As the material of coil was selected copper with conductivity $\sigma = 5.997 \times 10^7 \text{ S.m}^{-1}$. Coil was driven by DC current $I_0 = 20\text{A}$. Very important parameter of simulation is a mesh. Proper size and shape of mesh elements influences the accuracy of solution; the finer the mesh elements, the better the solution accuracy. Of course, very fine mesh can be a problem and computer cannot solve such model. It is very important to find optimal number of elements. In our case the triangular mesh of 288500 elements with maximum size of 2mm was created. Inductance of coil is then determined from total magnetic energy created by current passing through the coil using the following formula.

$$W_m = \frac{1}{2} L I_0^2 \quad (7)$$

The problem was solved as a magneto-static problem and calculation took 20s. Total magnetic energy is $W_m = 10.534 \times 10^{-4} \text{ J}$.

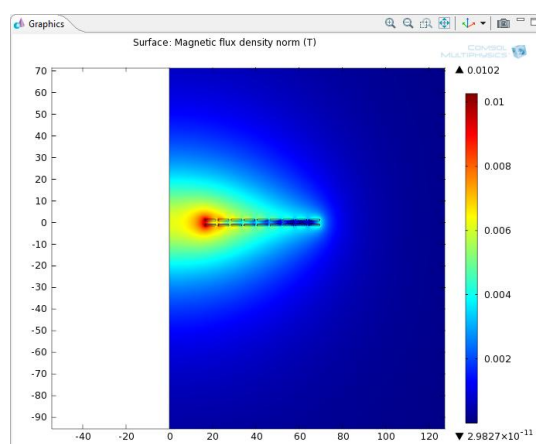


Fig. 5. Magnetic flux density [T].

Inductance of coil using (7) is $L = 24.04\mu\text{H}$. In Fig. 5 is shown magnetic flux density. The magnetic field is strongest in the middle of the coil and decreases with increasing radius of the coil.

VII. CONCLUSION

The aim of the paper was to estimate of inductance of double-layer spiral coil using finite elements method of solution and comparing the results with analytic solution. First, in analytic approach, we had to calculate inductance of single-layer coil and then using the equation for mutual

inductance could be calculated the inductance of double-layer spiral coil. For double-layer coil using software package we can calculate the inductance directly.

The software solved the total magnetic energy created by the coil and using simple formula we can calculate inductance of coil. The difference between results of analytic and simulation solution is 4.74%.

The finite element method of solution can be used for estimating other important properties of coil such as: self-capacity or resistivity of the coil.

The future work will be focused on these properties and comparing them with experimental results.

ACKNOWLEDGMENT

The paper has been prepared under support of Slovak grant projects KEGA No. 005TUKE-4/2012, KEGA No. 024TUKE-4/2012 and VEGA No. 1/0559/12.

REFERENCES

- [1] M. Pospisilik, L. Kouril, I. Motyl, M. Adamek, "Single and Double Layer Spiral Planar Inductors Optimisation with the Aid of Self-Organising Migrating Algorithm," Recent Advances in Signal Processing, Computational Geometry and Systems Theory, Proceedings of the 11th WSEAS International Conference on Signal Processing, Computational Geometry and Artificial Vision (ISCGAV '11) Proceedings of the 11th WSEAS International Conference on Systems Theory and Scientific Computation (ISTASC '11), Florence Italy, August 23-25, 2011, pp. 272-277.
- [2] E. Paese, M. Geier, J. L. Pacheco, R. P. Homrich, J. C. S. Ortiz, "Mathematical modeling of an electromagnetic forming system with flat spiral coils as actuator," The 4th International Conference of High Speed forming, 2010, pp. 219-228.
- [3] W. Tang, Y. Zhu, Y.L. Chow, "Inductance calculation of spiral inductors in different shapes," Microwave Conference Proceedings, 2005. APMC 2005. Asia-Pacific Conference Proceedings, vol.5, no., pp. 3 pp., 4-7 Dec. 2005
- [4] J. Zhao, "A new calculation for designing multilayer planar spiral inductors, In Electronics Design, Strategy, News, July 2010
- [5] S.S. Mohan, M. del Mar Hershenson, S.P. Boyd, T.H. Lee, "Simple accurate expressions for planar spiral inductances," Solid-State Circuits, IEEE Journal of, vol.34, no.10, pp.1419-1424, Oct 1999
- [6] I. Kováčová, D. Kováč, T. Vince, "Electromagnetic compatibility", Published by TU FEI, Košice, pages 138, 2009, ISBN 978-80-553-0150-1
- [7] D. Kováč, I. Kováčová, J. Molnár, "Elektromagnetic compatibility – measurement", Published by TU, FEI, pp. 72, 2009, ISBN 978-80-553-0151-8
- [8] I. Kováčová, D. Kováč, D. "Industrial Electrical Engineering – Modeling of Converters", 1st part, book, Published by TU of Košice, 2011, 66 pages, ISBN 978-80-553-0617-9
- [9] I. Kováčová, D. Kováč, D. "Industrial Electrical Engineering – Modeling of Converters", 2nd part, book, Publisher TU of Košice, 2011, 68 pages, ISBN 978-80-553-0618-6

DAC Testing by using a precise Reference DC Voltage and Dithering Signal

¹Martin SEKERÁK (3st year), ¹Marek Godla (2nd year), ²Jozef LIPTAK (1st year),
Supervisor: ¹Linus MICHAELI, ²Jan SALIGA

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

martin.sekerak@tuke.sk, marek.godla@tuke.sk, jozef.liptak@tuke.sk

Abstract— The precise Digital to Analog Converter (DAC) is one of the main parts of many electronic equipment and their non-linearity and distortion directly influence properties of the generated signal. The improvements in DAC performance together with increasing resolution is challenging problem for testing. Proportionally to both trends the testing accuracy and speed should increase. Proposed method provides DAC linearity testing under dynamic condition by comparison of the DAC output voltage with the reference signal using fast comparator and strobed latch registering DAC input code at the equality of voltages. The reference signal is the superposition of DC voltage and the dithering signal. Accuracy of DC measurement together with the linearity of dithering signal is responsible for the metrological traceability. The proposed data processing algorithm determine Integral non-linearity (INL) and Differential non-linearity (DNL) from recorded code words. Theoretical analysis and simulation included in paper affirm the performance of the proposed technique.

Keywords— Digital to analog converter; DAC; DAC dynamic test; INL; DNL; Dithering, Comparator

I. INTRODUCTION

DAC converters are key components of precise synthesizers. The use of such DACs requires a simple and reliable testing procedure realizable in the laboratories equipped with general purpose instruments. Many DAC procedures have been proposed by different authors [1] - [25]. Some of them have been assessed both by simulation and experimental implementation; however the implementation of the proposed methods is limited by the hardware restrictions. Some of them require spectral analyzers with accuracy higher than the nonlinearities of the DAC under test.

The authors proposed in a previous paper [13] a dynamic DAC testing procedure based on the conversion of the DAC output level into a time interval. The advantage in comparison to the previous proposals and standardized methods was the transfer of precision requirements from the waveform recorders to the time interval measurement. The time interval is measurable with higher precision than the dynamic output voltage. The drawback of this method was that besides the comparator accuracy the time jitter of the clock frequency corrupted the final accuracy. In order to avoid errors caused by using counters for measurement of short time intervals the presented method is proposed. It avoids counters by direct registration of code levels in a shift register.

Proposed method is based on the comparison of the dynamic output voltage of the DAC under test (DAC UT) with the reference which is superposition of the generated DC voltage and dithering voltage by using a comparator. DAC UT is excited by the triangular digital signal, where signal slope is adjustable. The average DC value of the reference is measured by the precise DC voltmeter determining testing accuracy. The moment when the DAC analog output voltage exceeds the reference level serves for registration of the corresponding digital input code. The output of the comparator is used as the control signal to register the digital input code word of the DAC under test in a fast memory. This approach allows assigning to the DAC input digital code the precise (real) value of the DAC output voltage which is utilized as the reference voltage. Digital post-processing for the same reference allows determining relation between input digital code and DAC output voltage. Performing of this procedure along DAC full scale range allows to determine INL(k) and DNL(k) as basic error criteria.

The main advantages of the proposed dynamic methods compare to alternative DAC dynamic test methods is using predominantly digital signal processing to determinate functional DAC characteristic, using high precision DC voltmeter which are relatively price available, the insensitivity to frequency jitter and possibility to simultaneously measure a few neighboring transient code levels from only one record within one test run and thus effectively reduce the test run time needed to measure full transfer characteristics.

II. THE PROPOSED METHOD FOR TESTING DAC

The digital sawtooth sequence represents exciting digital signal for DAC UT. The output voltage of the DAC UT impacted by nonlinearity errors is compared with the DC reference voltage by the fast voltage comparator. The output of the comparator is the control signal to register the digital input code word of the DAC under test into a fast memory. This approach allows assigning to the DAC input digital code the precise (real) value of the DAC output voltage which is utilized as the reference voltage. In order to avoid quantization uncertainty the triangular dithering voltage with known peak-peak value is superimposed to the reference voltage ($U_{SUM} = U_{DC} + U_{DITH}$). The required amplitude of dithering is achieved by using an appropriate resistive divider. The impact of the dithering voltage is suppressed due to the averaging effect

during an integer number of dithering periods. The distortion of triangular dithering voltage can involve uncertainty. The amplitude reduction by the passive divider decreases the existing distortion of the generated dithering voltage at the output of the dithering DAC. Moreover the amplitude reduction allows the use the dithering DAC with a lower resolution than the resolution of the DAC under test. The output value from the precise DC voltmeter during one testing interval determines reference voltage with required accuracy. In such a case the requirement on the precision of determination of the DC voltage is granted by using a precision voltmeter and no special instruments like reference DC source, signal analyzers with accuracy higher than DAC UT are required.

The offset error and time delay of the comparator are being reduced by the digital post processing. INL error at the beginning and at the end of DAC UT full scale range should be according the terminal definition equal to zero. The only consideration is that time delay is constant or is changing linearly along full scale range.

In order to avoid errors related with comparator's oscillation caused by the glitches on DAC UT output and comparator's settling time the output of the comparator is registered after time delay Δt .

A block diagram of the measurement stand for proposed method is shown in Fig. 1.

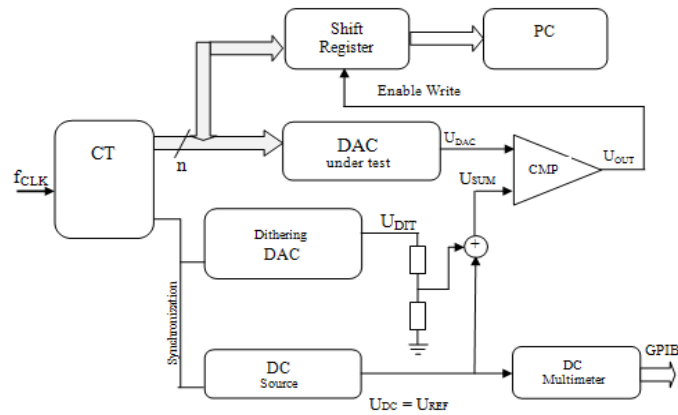


Fig. 1 Block diagram of the proposed measurement system.

One measuring cycle contains L periods of the sawtooth voltage generated by the DAC under test and M periods of the dithering voltage, as shown in Fig. 2. The peak to peak value of the dithering voltage W corresponds to a few LSBs of the DAC under test. The DC voltage measured by the precise DC voltmeter during this measuring cycle averages the reference voltage and determines only the precise value of the voltage from DC source.

The sawtooth control code $k(i)$ at the input of the DACUT in the i -th period is generated linearly by the instrumentation in each clock tact T_{CLK} by 1. The clock tact T_{CLK} is related to the period T_i as follows:

$$T_1 = 2^N * T_{CLK} \quad (1)$$

The period of generated analog sawtooth voltage T_1 can be adjusted by changing clock frequency f_{CLK} and in such a way the slope of the tested signal can be changed

$$S = \frac{U_{FS}}{2^N} * f_{CLK} \quad (2)$$

In order to avoid coherence between dithering voltage U_{DITH} and sawtooth voltage the constant L and M must be ordinary prime numbers. To meet the described requirement, we can write the following statement for periods of both signals and also the constant L and M :

$$\frac{T_2}{T_1} = \frac{L}{M} = \frac{f_1}{f_2} \quad (3)$$

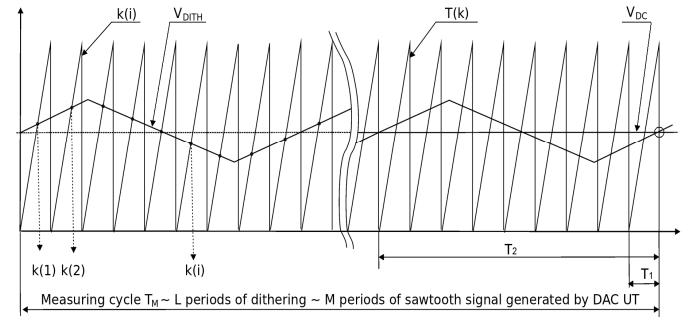


Fig. 2 Analog output voltage $U_{DAC}(k)$, from the DAC under test and reference voltage U_{REF} as superposition of the dithering voltage U_{DITH} on DC voltage U_a .

During each sawtooth period the comparator detects the instant when the output voltage $T(k(i))$ of the DAC UT exceeds for the first time the voltage U_{SUM} , which is the sum of the DC voltage (U_{DC}) and the dithering voltage (U_{DITH}). This time instant is determined by a positive derivation of the output voltage of the comparator U_{OUT} . We can write the following condition for this instant:

$$U_{DAC}(k(i)) \geq U_{SUM} = U_{DC} + U_{DITH} \quad (4)$$

The registered code word $k(i)$ from input of the DAC under test recorded in the shift register meets the following conditions:

$$T(k(i)) \geq U_{SUM} > T(k(i) - 1) \quad (5)$$

During one measuring period T_M the shift register stores an array of M values $k(i)$, $i = 1 \dots L$. Those values code words $k(i)$ are directly influenced by the difference between $U_{DAC}(k)$ and voltage U_{DC} .

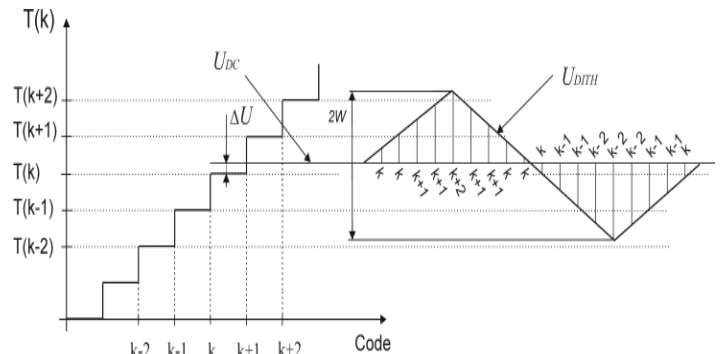


Fig. 3 Transfer characteristic of DAC UT together with reference voltage U_{DC} and dithering voltage U_{DITH} as well as the way of registration an array of M code words $k(i)$ in the shift register.

Tested transition code levels $T(k)$ for any k from the registered array is being transformed in the new array $p_k(i)$

$$p_k(i) = \begin{cases} 1 & \text{if } k(i) \leq k \\ -1 & \text{if } k(i) > k \end{cases} \quad (6)$$

This transformation serves mainly to suppress the

influence of neighboring quantization levels errors and nonlinearities dithering voltage on the overall accuracy of measurement. In addition makes the measurement almost completely independent of the amplitude and period of used dithering, which considerably simplifies the measuring conditions. The voltage difference $\Delta U(k)$ between tested transient code level $T(k)$ and analog voltage U_{DC} is then calculated as follows:

$$\Delta U(k) = W \cdot P_{AV} = W \left(\frac{1}{M} \sum_{i=1}^M p_k(i) \right) \quad (7)$$

Here P_{AV} represents average value of the array $p_k(i)$. The tested transient code level $T(k)$ at the output voltage of the DAC $U_{DAC}(k)$ for code k can be determined as:

$$T(k) = U_{DC} - \Delta U(k) \quad (8)$$

The integral nonlinearity of the tested code k than can be determined by following formula:

$$INL(k) = \frac{T(k) - k \cdot Q}{Q} \quad (9)$$

Calculation of two amplitude differences $\Delta U(k)$, $\Delta U(k-1)$ for two neighboring codes k , $k-1$ and the same reference voltage U_{DC} allows direct calculation of $DNL(k)$ from the same memorized array $k(i)$ using formula:

$$DNL(k) = \frac{\Delta U(k+1) - \Delta U(k) - Q}{Q} \quad (10)$$

Increasing dithering amplitude W allow speeds up whole testing process by estimating more transient code levels from one array $k(i)$ which was acquired for known value U_{DC} . On the other hand the increasing amplitude W increases the uncertainty of the determination ΔU . Uncertainty of the determination of W and its distortion is proportional to those parameters of the dithering DAC reduced by the division factor.

The total uncertainty of proposed method is given by the uncertainty $u(T(k))$ of determination of transition code level $T(k)$ (8). Error sources and uncertainties influencing the accuracy of measurement are:

- Errors and uncertainties of reference voltage U_{DC} that include errors and uncertainties of DC source and precise DC voltmeter
- Voltage comparator offset
- Non-ideality of dithering voltage covering its quantisation noise and nonlinearity of Dithering-DAC
- Uncertainty of finite record $p_k(i)$ in (6) given by statistical uncertainty and uncertainty of dithering amplitude W
- Errors caused by dynamic properties of the voltage comparator CMP. The overdrive dispersion and the dynamic errors of CMP can be suppressed by a convenient time shift Δt .

Because of different sources we can suppose these uncertainties are statistically independent.

III. VALIDATION RESULTS FROM SIMULATIONS

The proposed method has been simulated by the DAC with modelled non-linearities. The main goal of the simulation program was to verify theoretical analysis as well as confirm the impact of using various dithering amplitude and length of the recorded array on the final uncertainty. In this section selected dependencies will be presented from the obtained results. The first verified parameter with the most significant influence on resulting uncertainty of measurement was length of record M . The dependence of resulting uncertainty on record length M for constant amplitude W is shown in Fig. 4 and it is obvious that record length increased above 10 000 samples have only negligible influence on accuracy of achieved results.

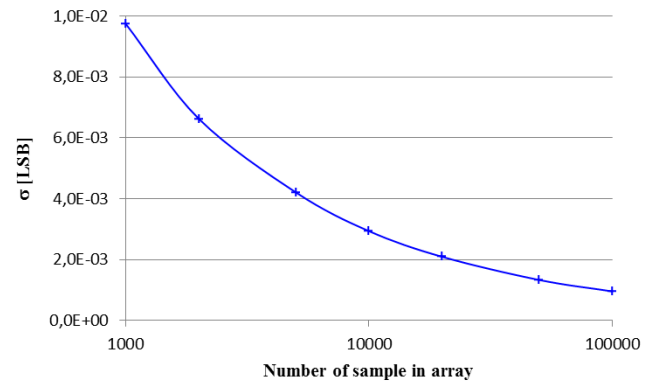


Fig. 4. Dependency of final uncertainty of evaluation INL characteristics on record length M for $W = 4$ LSB and $M/L = 11.48$.

The second verified influence of increased amplitude W on final uncertainty of measurement is shown on Fig. 5 as 3D graph depict this dependency for different record length M . Simulation results shows that uncertainty of measurement by proposed method slightly grows with increasing amplitude W , nevertheless the influence of amplitude of dithering is not as significant as influence of length of record M .

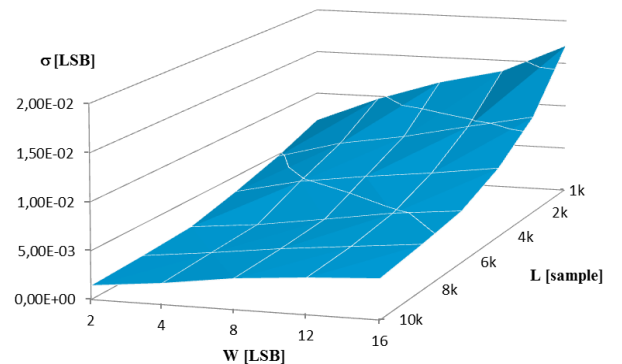


Fig. 5. Dependency of standard deviation of difference between INL characteristics on amplitude of dithering W for different record length M and constant ratio $M/L = 11.48$.

IV. CONCLUSION

In the paper, a new method for design and implementation of the measurement stand for DAC performance assessment has been presented. The method is

primary addressed to the characterization of high resolution DACs under dynamic condition. It is based on the comparison of the dynamic output voltage of the DAC under test with the reference DC voltage by using a comparator. Approach has been verified by created simulation of proposed method to verified impact of various working conditions as well as the influence of individual parameters on resulting uncertainty of evaluation modeled error. The advantage of the proposed method is based on the determination of the functional DAC parameters (INL and DNL) in dynamic mode using predominantly digital signal processing of registered data and using affordable digital voltmeter as metrological traceability. There is also a few ways as to improve properties of proposed method or speed to become more interesting for real measurement, however it will be presented in other publication.

Study of the simulation results of proposed method under various working condition confirmed theoretical analysis that length of record array M is the most significant factor influencing resulting uncertainty. The uncertainty decrease with increasing number of sample M in one measurement cycle, however increasing M extends the measuring time. Application of the proposed method requires choosing a compromise between required uncertainties and measuring time according to user's needs. Amplitude of dithering W has as well influence on final uncertainty; nevertheless it is significantly lower than influence of the record length of array and moreover with increasing M decreases the influence of all other components of uncertainty.

ACKNOWLEDGMENT

The work is a part of project supported by the Science Grant Agency of Slovak republic (No. 1/0555/11).

This work was supported by Agency of the Ministry of Education of the Slovak Republic for the Structural Funds of the EU under the project Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020).

REFERENCES

- [1] D.L. Cari, D. Grimaldi, "Static characterization of high resolution DAC based on over sampling and low resolution ADC", Proc. Of IEEE Instrum. And Measur. Techn. Conf., IMTC 2007, May 1-3, 2007, Warsaw Poland
- [2] C. W. Lin, S. F. Lin, and S. F. Luo, "A new approach for nonlinearity test of high speed DAC," in Proc. IEEE Int. Workshop Mixed-Signals, Sens., Syst., 2008, pp. 1-5.
- [3] Baccigalupi, A., Liccardo, A., Grimaldi, D., Carni, D.L.: "Digital to analog converters test based on time to voltage conversion", Proc. of IEEE Instrumentation and Measurement Technology Conference, Binjiang, Hangzhou; May 2011, Article number 5944277, pp. 6-11
- [4] F. Xu, "Signal cancellation technique for testing high-end digital-to-analog converters," in Proc. IEEE IMTC, 2007, pp. 1-4.
- [5] A. Baccigalupi, M. D'Arco, A. Liccardo, M. Vadursi, "Test equipment for DAC's performance assessment: design and characterization," accepted for publication on IEEE Trans. on Instrumentation and Measurement.
- [6] M. D'Arco, A. Liccardo, and M. Vadursi, "Design of a test equipment for DAC's performance assessment," in Proc. IEEE I2MTC, Singapore, May 5-7, 2009, pp. 1232-1237.
- [7] D.L. Carni, D. Grimaldi, "Static and Dynamic Test of High Resolution DAC Based on Over Sampling and Low Resolution ADC", Measurement, vol. 43, No. 2, February 2010, pp. 262-273.
- [8] L. Angrisani and M. D'Arco, "Modeling timing jitter effects in digital-to-analog converters," IEEE Trans. Instrum. Meas., vol. 58, no. 2, pp. 330-336, Feb. 2009, Digital Object Identifier 10.1109/TIM.2008.2003313.
- [9] D.L. Carni, D. Grimaldi, "Characterization of High Resolution DAC by DFT and Sine Fitting", Proc. of IEEE I2MTC, Singapore, May 5-7, 2009.
- [10] L. Jin, H. Haggag, R. Geiger, and D. Chen, "Testing of precision DACs using low-resolution ADCs with wobbling," IEEE Trans. Instrum. Meas., vol. 57, no. 5, pp. 940-946, May 2008.
- [11] E. Balestrieri, P. Daponte, S. Rapuano, "Digital to analogue converters: a metrological overview," in Proc. of IMEKO TC-4, 29 Sept.-1 Oct, 2004, vol.2, pp.875-880.
- [12] Draft Standard for Terminology and Test Methods for Digital-to-Analog Converters, IEEE STD P1658, Sep. 2008
- [13] M. Sekerák, Michaeli, L., Šaliga, J., A. Cruz Serra, "Methods with a new approach for measure static characterization of High Resolution DAC converters in Proc. of 17th Symposium IMEKO TC 4, TC 19 and Sept. 8-10, 2010, Kosice, Slovakia.
- [14] D.L. Carni, D. Grimaldi, "Comparative analysis of different acquisition techniques applied to static and dynamic characterization of high resolution DAC", Proc. of XIX IMEKO World Congress "Fundamental and Applied Metrology", Lisbon, Portugal, September 6-11, 2009.
- [15] A. Baccigalupi, M. D'Arco, A. Liccardo, M. Vadursi, "Testing high resolution DACs: a contribution to draft standard IEEE P1658", Measurement, Vol. 44, Issue 6, July 2011, pp. 1044-1052.
- [16] J. Savoj, Ali-Azam Abbasfar, A. Amirkhany, B.W. Garlepp, M.A. Horowitz, "A new technique for characterization of Digital-to-Analog Converters in high-speed systems" Design, Automation & Test in Europe Conf. & Exhibition, 16-20 April 2007, pp.1-6.
- [17] B. Vargha, J. Schoukens, and Y. Rolain, "Static nonlinearity testing of digital-to-analog converters," IEEE Trans. Instrum. Meas., vol. 50, no. 5, pp. 1283-1288, Oct. 2001.
- [18] Adamo, F., Andria, G., Attivissimo, F., Giaquinto, N.: Linearisation of A/D converters by dither and Chebyshev polynomials, /In: Measurement. ISSN 0263-2241, Vol. 35, no. 2 (2004), pp. 123-130.
- [19] Kim, I., Jang, J., Son, H., Kang, S.: "A new static test of a DAC with a built-in structure", Proc. of Midwest Symposium on Circuits and Systems, Seoul, 2011, ISBN: 978-161284857-0, Article number 6026361
- [20] Huang, X.-L., Huang, J.-L.: "ADC/DAC loopback linearity testing by DAC output offsetting and scaling", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Volume 19, Issue 10, October 2011, Article number 5559497, pp. 1765-1774
- [21] Demidenko, S., Mohtar, A.Z., Lee, K.H.: "Microcontroller testing using on-load-board DAC", Proc. of IEEE Instrumentation and Measurement Technology Conference, Binjiang, Hangzhou; May 2011, Article number 5944348, pp. 12-15
- [22] Di Nisio, A., Lanzolla, A.M.L., Savino, M.: "Simultaneous A/D and D/A converters linearity testing with deterministic dithering", Proc. of IEEE Instrumentation and Measurement Technology Conference, Binjiang, Hangzhou; May 2011, Article number 5944236, pp. 208-212
- [23] D'Arco, M., Liccardo, A., Pasquino, N.: "Evaluating DACs linearity and intermodulation errors through an ANOVA approach", Proc. of IEEE Instrumentation and Measurement Technology Conference, Binjiang, Hangzhou; May 2011, Article number 5944196, pp. 1040-1044
- [24] Fexa, P., Vedral, J., Svatoš, J.: "Dac testing using Modulated Signals", Metrology and Measurement Systems, Volume 18, Issue 2, 2011, Page 10
- [25] W. Kester, "Evaluating high speed DAC performance", Analog Device, MT-013, January 2006, available at www.analog.com.

- [26] H. Xing, D. Chen, and R. Geiger. On-chip at-speed linearity testing of high-resolution high-speed DACs using DDEM ADCs with dithering. In *Electro/Information Technology*, pages 117-122, 2008.

Dielectric Spectroscopy of XLPE Cables

¹Martin GERMAN-SOBEK (1st year), ²Jozef KIRÁLY (1st year)
Supervisor: ³Roman CIMBALA

¹Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

¹martin.german-sobek@tuke.sk, ²jozef.kiraly@tuke.sk, ³roman.cimbala@tuke.sk

Abstract—Cross-linked polyethylene (XLPE) has become the world's preferred modern insulation for high voltage cables. The article deals with the measurement of current response of XLPE cable specimens using KEITHLEY 617 electrometer. The measurement was carried out by the method of time domain dielectric relaxation spectroscopy.

Keywords—XLPE cable, dielectric spectroscopy, destruction free diagnostic, cable insulation.

I. INTRODUCTION

In the present, the increased demand for electricity is dominated. From electricity view, this means that for high-voltage electrical equipments is required greater security and reliability. These requirements relate mainly to the insulation system, which is an important and sensitive part of any electrical equipment.

Cross-linked polyethylene (XLPE) is now widely used as electrical insulation material for low and high voltage distribution cables because of its excellent physical, chemical and electrical properties. Given that these cables may be exposed to several kA current and hundreds of kV voltage, they are critical parts of the transmission infrastructure and is expected their high resistance against possible failures during their lifetime [10].

During operation, the cable is permanently exposed to thermal aging, which may cause irreversible damage of cable insulation. Voltage, mechanical, chemical and thermal stress and pollution of environment are the primary initiators for the degradation of insulation. Process of aging of insulation is the most acting on the quality of insulation and it is a phenomenon that is essentially can not be affected. Therefore, it is necessary to assess degradation and insulating state of cables.

There are several methods for detection of insulating state of XLPE insulation. One of the methods is the dielectric relaxation spectroscopy (DRS).

II. DIELECTRIC RELAXATION SPECTROSCOPY

This method is one of the non-destructive measurement methods and it is based on a wide range of measurements in time and in the frequency domain. Generally speaking, the behaviour of the dielectric material is characterized DC conductivity, dielectric response function and high-frequency component of the relative permittivity [3].

This article deals only with the measurement of time

response. When evaluating the time response is usually used the polarization and depolarization current for evaluate of system.

Ideal dielectric is a substance that contains electric charges bound by electrostatic forces. In the case of real dielectrics, there are also free charge carriers, which create undesirable conductivity.

Dielectric polarization is a phenomenon that occurs under the effect of electric field and causes the movement of free charge carriers [7]

If dielectric material is inserted in the electric field of intensity $E(t)$, this material will be polarized. Electric polarization $P(t)$ in dielectric materials can be divided into two parts, namely the fast and slow polarization [3].

Slow polarization processes are measured by charging and discharging currents.

Current $i(t)$, which flows through the dielectric when exposed to electric field of intensity $E(t)$ can be expressed as:

$$i(t) = C_0 \left[\frac{\sigma_0}{\epsilon_0} u(t) + \epsilon_\infty \frac{du(t)}{dt} + \frac{d}{dt} \int_0^t f(t-\tau) u(\tau) d\tau \right] \quad (1)$$

where:

C_0 - geometric capacity of the measured object

Polarizing current $i_{pol}(t)$ can be expressed as:

$$i_{pol}(t) = C_0 U_c \left[\frac{\sigma_0}{\epsilon_0} + f(t) \right] \quad (2)$$

As a result of a DC voltage U_c , polarizing current is composed of two parts. The first part is related to conductivity of the investigated object and the second part is related to activation of different polarization processes inside the investigated material [3, 7].

If the DC voltage U_c is removed at the time T and the material is grounded, depolarization current $i_{depol}(t)$ starts to flow through the material.

$$i_{depol}(t) = -C_0 U_c [f(t) - f(t + T_c)] \quad (3)$$

where:

$f(t)$ - monotonically decreasing function [3]

A course of current response after connecting and disconnecting the DC voltage (charging and discharging current) is shown in Figure 1.

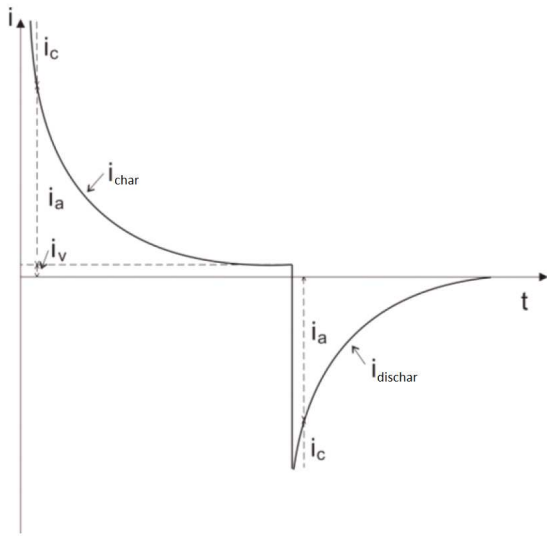


Fig. 1. Current response after connecting and disconnecting the DC voltage [7]

The second part of equation (3) can be neglected for large values of TC. Depolarization current becomes proportional to the dielectric response [3]:

$$f(t) \approx -\frac{i_{depol}}{C_0 U_c} \quad (4)$$

The total polarizing current flowing through the dielectric can be following assumption of existence the independent Debye's polarization processes and Maxwell-Wagner equivalent model, expressed as the sum of currents with exponentially decreasing amplitude with definite-time component, which represents the individual of the polarization occurring in the material by the equation:

$$i(t) = \frac{U}{R_0} + \sum_{i=1}^n I_{mi} \exp\left(\frac{-t}{\tau_i}\right) \quad (5)$$

where:

U – applied DC voltage

R_0 – DC insulation resistance after an infinitely long time

I_{mi} – amplitude of i -th component of Debye's elementary current

τ_i – relaxation time constant of the i -th component [7]

Number of members of expansion n is dependent on the structure and homogeneity of the dielectric and the time at which the polarizing action are measured [7].

III. EXPERIMENT

Diagnostic measurement was performed by the method of dielectric relaxation spectroscopy using Keithley 617 electrometer for two specimens of HV cable:

- specimen A – operationally unaged, undamaged cable,
- specimen B – cable of operationally unknown technical condition.

Circuit configuration for measurement is shown in Figure 2, where C_X is measured specimen of cable.

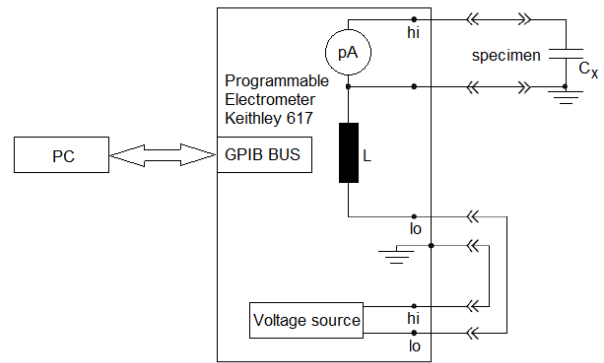


Fig. 2. Schematic diagram of measuring circuit

For both cables, it was a power cable with an aluminium core and XLPE insulation. Specimens of cables were approximately 1 m long and protective cable jacket and semiconducting layer have been removed (10 cm at both ends) and shielding taken out. Time of measurement was 1000 s.

Aim measuring has been comparison of measured dependence of charging current of two different operationally aged specimens of cable.

Figure 3 and Figure 4 shows the time dependence of charging current operationally unaged specimen A and operationally unknown condition of specimen B, respectively.

The displayed waveforms have hardly a significant difference. For more visual comparison was from measured dependence of charging currents calculated constants of elementary currents and relaxation times for seven polarization processes. Calculated data are shown in Table 1 and Table 2.

For further interpretation was excluded the development of these amplitude of current and relaxation time, which included values less than or equal to 0 or greater than 10^7 , because it is not possible that currents were negative values.



Fig. 3. Current response of specimen A

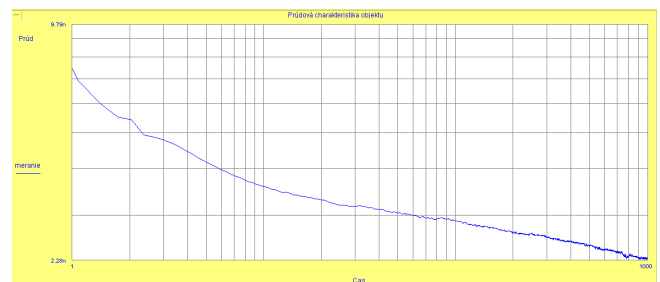


Fig. 4. Current response of specimen B

TABLE I
CONSTANTS OF ELEMENTARY CURRENTS AND RELAXATION TIMES FOR
SEVEN POLARIZING PROCESSES – SPECIMEN A

Approximation	Relaxation time τ_i (s)	Relaxation current I_{mi} (pA)
1	2723,85	3850
2	6,29 3072,3	3614 3751
3	1,83 131,52 5735,41	6513 947 3367
4	0,34 4,56 218,36 9799,66	34818 2587 1056 3089
5	0,27 3,34 28,01 290,08 16185,93	56705 2749 439 1021 3006
6	0,2 2,5 14,73 225,94 3044238552 33521844986	130328 3046 656 954 1041226562 -1041223413
7	0,2 2,5 14,73 225,94 3077452034 33535729306 1,73628E+11	1300325 3046 656 954 974264972 0 -974261823

TABLE II
CONSTANTS OF ELEMENTARY CURRENTS AND RELAXATION TIMES FOR
SEVEN POLARIZING PROCESSES – SPECIMEN B

Approximation	Relaxation time τ_i (s)	Relaxation current I_{mi} (pA)
1	2977,85	3013
2	4,14 3433,02	4907 2930
3	1,38 50,28 4575,89	8573 884 2792
4	0,31 4,38 93,82 5270,59	46506 2874 616 2732
5	0,28 3,69 48,68 632,74 920782843,3	56728 3042 557 807 2116
6	0,29 3,78 55,02 233086,86 56281044902 -6,50337E+11	55289 3023 582 42982590 -9,50087E+12 9,50083E+12
7	0,29 3,78 55,02 233815,19 56281050582 33535729306 1,73628E+11	55284 3023 582 43251653 -9,53073E+12 0 -974261823

As shown in Table 1 and Table 2, values of time and current for the 6th and 7th approximation are incorrect, since they reach zero or negative values. For this reason, the values τ_i and I_{mi} for $n=5$ polarization processes were used in analyzing, thus the 5th approximation.

IV. CONCLUSION

The quality of insulation and her electrical parameters are dependent on a number of factors that affect the insulation during operation. The elementary polarization currents are dependent on the size of measured object and its capacity, therefore, for analyze the quality of materials is an important analysis of stabilization time constants of elementary currents. Comparing the time constants obtained from Table 1 and Table 2 for the 5th approximation shows that specimen B has a higher degree of degradation as specimen A, since time constant τ_{55} is significantly greater for specimen B. Therefore, it can be confirmed that the specimen B is operationally degraded. At the same time it can be stated that the use of constants to differentiation measured specimens and reliable determination of insulation state of material is appropriate to apply the method to examine and evaluate their properties as a whole. A suitable method is a neural network.

ACKNOWLEDGMENT

This publication is the result of the Project implementation: Research centre for efficient integration of the renewable energy sources, ITMS: 26220220064 supported by the Research & Development Operational Programme funded by the ERDF.



We support research activities in Slovakia / Project is cofinanced from EU funds.

REFERENCES

- [1] Marton, K.: *Elektrotechnické materiály (Fyzika dielektrík)*, ESVŠT, Košice, 1971.
- [2] Mizutani, T.: "Space charge measurement techniques and space charge in polyethylene," *IEEE Transaction on Dielectrics Electrical Insulation*, Vol. 1, No. 5, 1994, pp. 923-933.
- [3] Zaengl, W. S.: "Dielectric Spectroscopy in Time Domain for HPV Power Equipment, Part I: Theoretical Consideration," *IEEE Transaction on Dielectrics Electrical Insulation*, Vol. 19, No. 5, 2003, pp. 5-19.
- [4] Model 617 Programmable Electrometer, Instruction Manual, Keithley Instruments, Inc., Cleveland, Ohio, USA, 1984
- [5] Palai-Dany, T.: "Dielectric Relaxation Spectroscopy with Keithley 617 Electrometer, *EEICT*, CVUT, Brno, 2006.
- [6] Cimbal, R.: "Dielektrické vlastnosti XLPE káblov" In: *Starnutie Elektroizolačných systémov*. Vol. 1, No. 1, 2006, pp. 14-16.
- [7] Cimbal, R.: *Starnutie vysokonapäťových izolačných systémov*, TUKE, Košice, 2007, 188p.
- [8] Muhr, M., Schwarz, R., Jaufer, S.: "Electrical Measurements as Diagnostic Tool for HV-Insulations," Graz, 2005, pp. 1-11.
- [9] Hoff, G., Kranz, H. G.: "Interpretation of Dielectric Response Measurement Data from Service Aged XLPE-Cables," *IEEE 7th*

International Conference on Solid Dielectrics, Eindhoven, June 25-29, 2001, pp. 381-384.

- [10] Fothergrill, J. C., et al.: "The Measurement of Very Low Conductivity and Dielectric Loss in XLPE Cables," *IEE Transaction on Dielectrics Electrical Insulation*, Vol. 15, No. 5, 2011, pp. 1544-1553.

Influence of concentration of magnetic nanoparticles in transformer oil based ferrofluid from view of dielectric spectroscopy

¹Jozef Király (1st year), ²Martin German-Sobek (1st year)
Supervisor: ³Roman Cimbala

^{1,2,3}Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic

¹jozef.kiraly@tuke.sk, ²martin.german-sobek@tuke.sk, ³roman.cimbala@tuke.sk

Abstract - The article deals with comparison of current-time characteristics measured on magnetic fluids based on inhibited transformer oil with different concentration of nanoparticles Fe₃O₄ and on pure transformer oil.

Keywords—ferrofluid, nanoparticles, current-time dependences, transformer oil

I. FERROFLUIDS IN GENERAL

Ferrofluid is liquid, which consist from carrier fluid and ferromagnetic or ferromagnetic nanoparticles coated by surface active substance. Respect to size of particles ferrofluids are colloidal liquids. Common size of nanoparticles is from 1 to 20 nm.

As mentioned, these particles are coated by surface active substance – surfactant. In our case takes the role of surfactant oleic acid, which is monounsaturated omega-9 fatty acid. Molecules of oleic acid are made from polar hydrophilic head and long unipolar hydrophobic chain. Main function of surfactant is to prevent the nanoparticles from clumping together (see fig.1). Magnetic attraction between nanoparticles is weak enough that the surfactant's Van der Waals force is able to prevent from clustering. Particles which were created from many nanoparticles would be too heavy to stay dispersed in volume of liquid by Brownian motion. In case of usage magnetic fluid as replacement of transformer oil is important to note that the settled particles can cause inhomogeneity in power transformer insulation system.

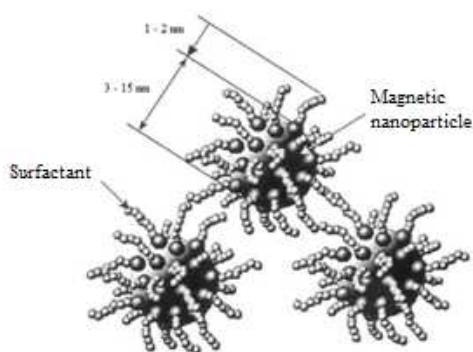


Fig. 1. Nanoparticle of Fe₃O₄ coated with oleic acid.[1]

II. THERMOMAGNETIC CONVECTION IN MAGNETIC FLUIDS AND USAGE IN POWER TRANSFORMER

Condition for this phenomenon is presence of temperature gradient and magnetic field. For use in cooling system of power transformer are conditions for thermomagnetic convection satisfied. The source of temperature gradient is heat produced by AC current in windings. Thus heat generated by winding and core of transformer creates in area inhomogeneous magnetic field temperature gradient. The velocity of transmission of heat from inside of power transformer to walls of the vessel and dissipation of heat to surrounding environment determines highest currents in the windings therefore in part determines a size and weight of power transformer for a given power. Accordingly more efficient cooling could bring smaller dimensions of power transformers and less amount of cooling fluid. Equally important is decreasing of windings temperature, which can be decreased up to 10%. Limit factor is a viscosity of these used magnetic fluids, because too dense magnetic fluid will not get into every layer of oil - paper insulation system. From view of electrical breakdown strength is also possible to think about positive impact magnetic nanoparticles. By degradation transformer oil begins releasing particles of sludge of microscopic size. These sludges get electrical charges what lead into formation of bridges, what cause electric breakdown. Presence of magnetic nanoparticles can partially reduce produce of these bridges, what should cause increasing of electric breakdown strength in partially degraded oils.[2]

III. DIELECTRIC SPECTROSCOPY

One of the non-destructive methods of diagnostics of high voltage insulation system in general is dielectric spectroscopy. This method is based on measurement of current in time and frequency domain. Stability of microstructure and composition of the insulating material change due to degradation processes. As a result of these changes is changing behavior of insulation material from view of polarization processes. Based on this fact, is possible characterize degree of degradation of specimen. Our sample can be mentioned as capacitor with pair of electrodes and

dielectrics between them. Thus after connecting of voltage to our specimen current starts flow thru dielectric. In case of ideal dielectric the current hasn't real component. But in practice is real component of current caused by processes of polarization and conductivity. These processes are dependent on kind of dielectric in which they are occurs. Based on existence independent Debye's polarization processes and Maxwell - Wagner equivalent model is possible to formulate total polarization current as:

$$i_c = \frac{U}{R_i} + \sum_{i=1}^n I_{m_i} \cdot e^{-\frac{t}{\tau_i}} \quad (1)$$

where:

U – applied DC voltage

R_0 – DC resistance after infinite time

I_{m_i} – amplitude of i – th component Debye's elementary current

τ_i – relaxation time constant i – th component [3]

Therefore overall polarization current is sum of currents with exponentially decreasing amplitude and with time independent component which represent. Number of members of development is dependent on the structure and homogeneity of the test dielectric and the time at which the polarizing action are measured [3].

IV. EXPERIMENT

A. Experimental base

For measuring of test specimens is necessary to prepare test bench which allows various setup of acting external stationary magnetic field. The external stationary magnetic field is provided by pair of NdFeB magnets with magnetic induction 40mT. Specimen was placed in vessel with volume approx. 100ml which is made from PTFE. For measurement of $i(t)$ characteristics we use electrode system consisting of pair of cylindrical electrodes with diameter 20 mm made from copper. The gap between electrodes was set to 1 mm. The configuration of vessel and other components is shown on fig. 2.

Specimen was made from carrier fluid - inhibited transformer oil ITO 100 and nanoparticles coated by oleic acid. The fluid was prepared by process of precipitation. The difference between our specimens was in concentration of nanoparticles contained in volume. The lowest used concentration of nanoparticles was 0,5% and highest concentration was 2%. The different concentration of specimens and three configurations of external magnetic fields provide good conditions for observing magnetic anisotropy. The first measurements were realized without using external magnets, thus vector B was equal to zero. In next two measurements were external magnets orientated in direction of electric field and perpendicular to the electric field. For comparison we measured also pure inhibited transformer oil ITO100.

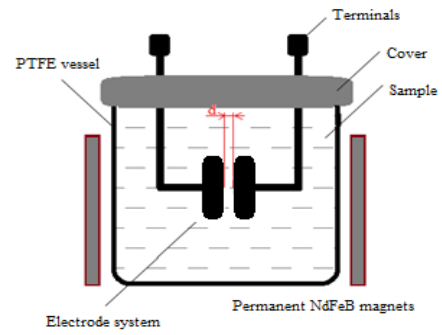


Fig. 2. Configuration of vessel with external magnets and electrode system.

The measurement was realized by measurement circuit which is shown on fig. 3. The metering instrument was used programmable electrometer Keithley 617 and the time of measurement was 1000 s. Acquired data was processed by Agilent HP VEE software.

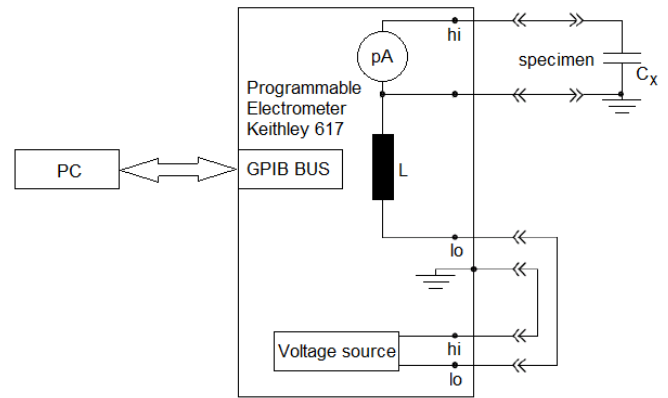


Fig. 3. Circuit configuration for measurement.

B. Measured characteristics

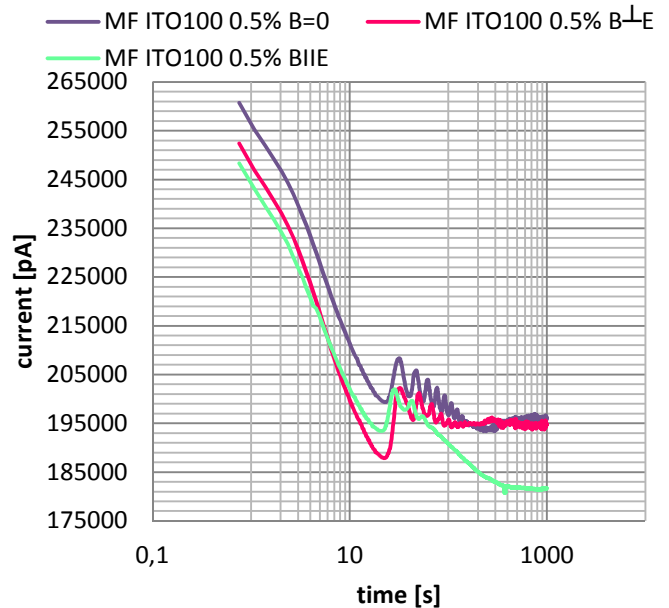
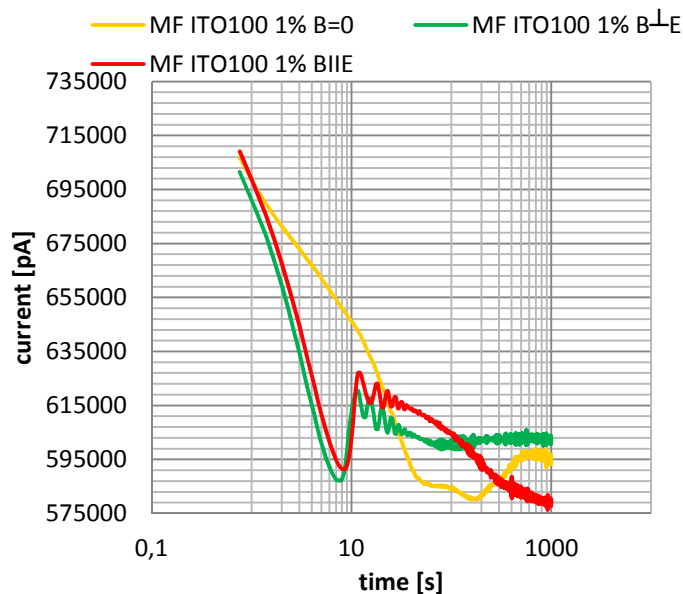
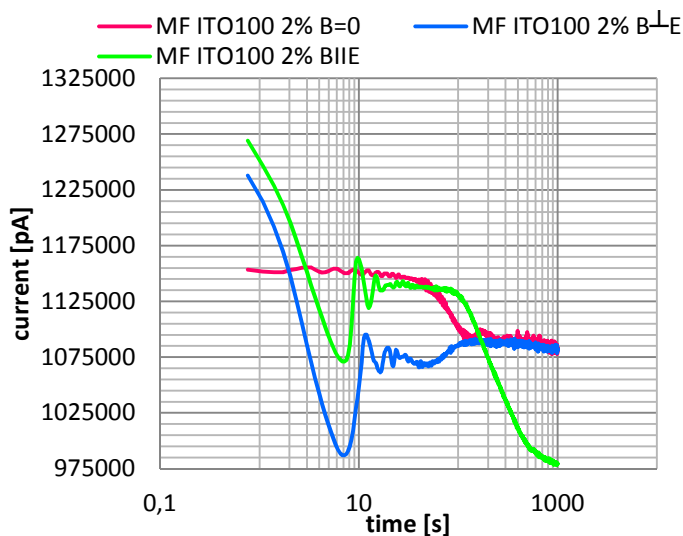
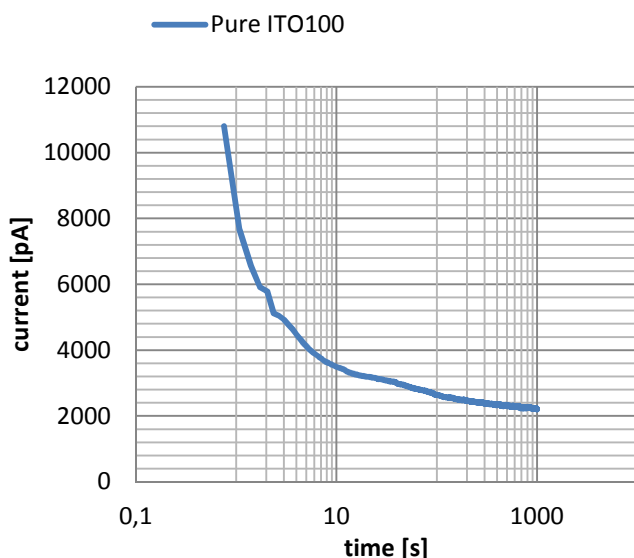
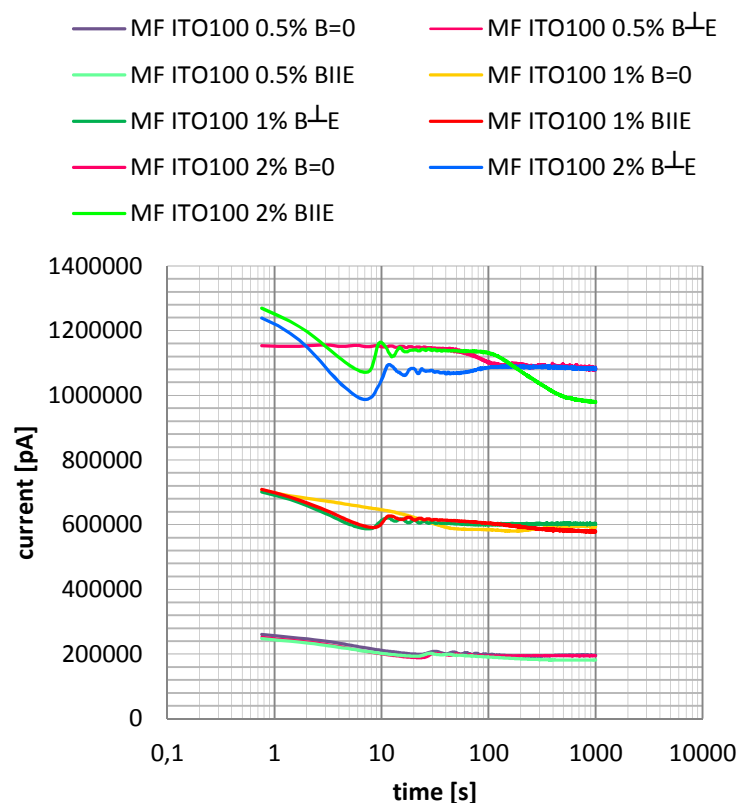


Fig. 4. $I(t)$ characteristics of MF ITO100 0,5%

Fig. 5. $I(t)$ characteristics of MF ITO100 1%Fig. 6. $I(t)$ characteristics of MF ITO100 2%Fig. 7. $I(t)$ characteristics of pure ITO100.Fig. 8. Comparison of $I(t)$ characteristics of MF ITO100 of different concentrations

V. CONCLUSION

We have to consider that measured current - time dependencies are macroscopic view of ongoing processes. Aim of this paper was to verify influence of magnetic nanoparticles concentration in a liquid volume of the ongoing processes on polarization. One reason is that a larger amount of magnetite nanoparticles in the magnetic fluid beneficial effect on the formation of long chains, which affect the polarization processes in the dielectric. Magnetic anisotropy which is shown on fig. 4 - 6 confirms the assumption, that action of external magnetic field in direction of electric field causes polarization current change. It's necessary to be mentioned that dependencies aren't affected only by the action of an external magnetic field but they are affected also with direction of these field. This fact is confirmed in every dependence. As is shown on fig. 7, pure transformer oil shows a much lower value of polarization current in whole time domain. Therefore is necessary to find out, how many nanoparticles can increase the thermal conductivity and breakdown strength, without negative impact to dielectric losses caused by processes of polarization.

ACKNOWLEDGMENT

This publication is the result of the Project implementation:
Research centre for efficient integration of the renewable
energy sources, ITMS: 26220220064 supported by the
Research & Development Operational Programme funded by
the ERDF.



We support research activities in Slovakia / Project is cofinanced from EU funds.

REFERENCES

- [1] MAYER, Daniel: Magnetické kapaliny a jejich použití (1. část). In: Elektro, roč. 7, 2007, č.3, p. 78-79.
- [2] HERCHL, František: Štúdium elektrofyzikálnych vlastností magnetických kvapalín na báze transformátorového oleja: Dizertačná práca. Košice SAV UEF, 2008. 87 p.
- [3] CIMBALA, Roman.: *Starnutie vysokonapäťových izolačných systémov*, TUKE, Košice, 2007, 188p.

Simulation of shadowing effect for UWB through wall surveillance in Comsol Multiphysics environment

Daniel URDZÍK

Supervisor: prof. Ing. Dušan Kocur, CSc.

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

daniel.urdzik@tuke.sk, dusan.kocur@tuke.sk

Abstract— Ultra wideband (UWB) radar appears to be the suitable technology for detection and tracking of people during rescue and law enforcement missions in harsh environments. However such a problem in detection and tracking of multiple targets may occur known as the shadowing effect. In this paper, the theoretical basis of the shadowing effect will be presented. To examine mutual shadowing in detail, the simulation in Comsol multiphysics program will be described. The results of the simulation will be presented. In the conclusion, the solution of this problem is briefly discussed.

Keywords— shadowing effect, shadows, shadow zones

I. INTRODUCTION

In recent years the use of short range UWB radar appears to be a suitable solution for the detection and localization purposes. The UWB radar system is the special type of radar and can be used to detect and track moving targets in harsh environments. One of the features of the UWB radar systems is their use for the through wall and through fire detection and tracking of moving targets during security and rescue operations. UWB radar emitted signals applied for these scenarios are at a relatively low frequency (typically between 100 MHz and 5 GHz). The result is that UWB radar has a fine resolution that also has an ability to penetrate many common building materials. Then, such devices are able to detect moving person by measuring changes in the impulse response of the environments. For the above outlined applications, short range UWB radar, with range up to 20-25m is often 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 [1], [2], [3]. The problem of multiple people tracking in real complex scenarios has been less well addressed. Our experiments gained from different measurements with the short range experimental UWB radar, which were dedicated to detection and tracking of multiple moving persons have shown that the detection of multiple persons is sometimes difficult. These difficulties are caused by the phenomenon of shadowing. The effect of shadowing have been previously described e.g. in [4], [5], [6], and [7] 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/absorption of emitted radar signals by the obstacles. However in the case of multiple person detection and tracking with UWB radar another type of shadowing occurs. 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 person shadowing will manifest itself. The quantitative description of the shadowing effect for the UWB detection and tracking applications have been described in [11], [12]. 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. As the consequence, 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 who 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 shadow zones with high additional attenuation where radar practically is usually not able to detect and track any of the persons which are located within these zones.

According to our best knowledge the problem of through wall surveillance in simulation interfaces have been analyzed in [8]. However in this work the author presents the simulation of the detectability of the human persons for the SAR applications. In this paper we will describe the simulations in Comsol Multiphysics simulation interface which were dedicated to examine the shadowing effect from the aspect of electromagnetic wave propagation theory. This paper is organized as follows. In section 2 the theoretical models of

shadowing effect will be described. In section 3 the physical model of shadowing effect will be analyzed. The last sections will be dedicated to the description of the simulation results and to the conclusions based on the results of the simulations.

II. THEORETICAL MODELS OF SHADOWING EFFECT

In Comsol Multiphysics simulation interface the theoretical geometric model of the person shadowing effect is needed to be defined for purposes of its description. It consists of a person located in a room with dimensions of 10x10m. This model computes the interaction between a human being and the incident electromagnetic field from a radar transmitter. The transmitter is considered to be distant enough that this field can be treated as a plane wave. This makes it possible to exclude the transmitter from the model geometry and look only at the person and its immediate surroundings. The geometrical model of the shadowing effect is shown in the Fig. 1.

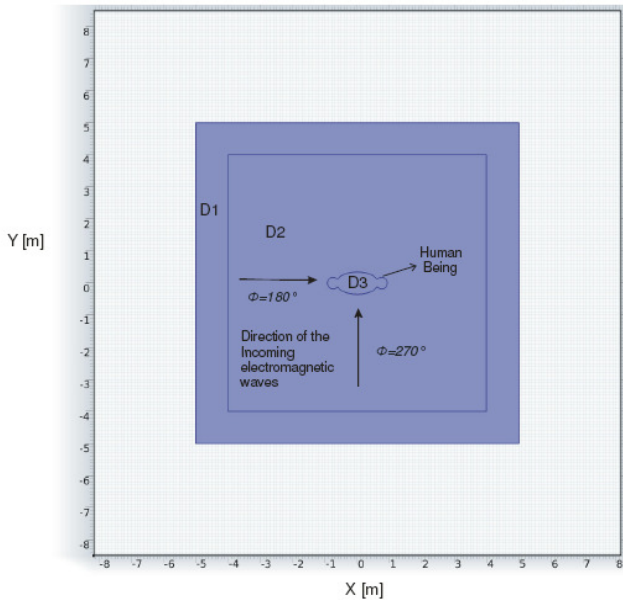


Fig. 1. Geometrical model of the shadowing effect

The 2D geometrical model can be divided into three spatial domains. In the centre the simplified model of human being is denoted as domain D3. The inner square in the geometry model represents air surrounding the human being, where the electromagnetic field will be calculated (domain D2). The outer square (domain D1) is a Perfectly Matched Layer (PML). This model shows the simplified situation, where the human being is exposed to the incoming electromagnetic waves produced by the radar.

III. PHYSICAL MODEL OF THE SHADOWING EFFECT

In Comsol Multiphysics simulation, the geometry model of the shadowing effect is needed to be associated with the physical model. The physical model describes the material properties of the environment, model of electromagnetic field calculation and the source of the electromagnetic field. Each domain in the geometric model in the Comsol Multiphysics simulation has to be associated with the proper equations which govern the physical problem that needs to be solved. The physical model of the shadowing effect can be divided

into three domains. In domains D2 and D3 the electromagnetic field is calculated. The domain D3 is perfectly matched layer (PML) domain. The PML minimizes unphysical reflections of the scattered electromagnetic wave as it leaves the model domain.

The background electromagnetic field emitted by the radar in the domains D2 and D3 is described by its out-of-plane electric field component [9]:

$$E_b = e^{(jk_0(x \cos(\phi) + y \sin(\phi)))} \quad (1)$$

In this equation:

- j is the imaginary unit,
- $k_0 = 2\pi f/c$ is the wave vector in vacuum,
- c is the speed of light,
- f is the frequency
- ϕ is the angle of incidence of the radiated electromagnetic field.

By changing the parameter ϕ we can change the direction of electromagnetic field produced by the radar with respect to the position of the person.

The time harmonic wave equation is solved for the relative field, $E_{rel} = E - E_b$, where E is the total measurable field. Physically, the relative field describes the difference to the measured field caused by the presence of the person. It is this difference that defines how detectable the person to the radar. The electric and magnetic component of the electromagnetic field is computed by the following equation [10]:

$$\nabla \times (\mu_r^{-1} \nabla \times E_{rel}) - (\epsilon_r - j \frac{\sigma}{\omega \epsilon_0}) k_0^2 E_{rel} = 0 \quad (2)$$

In order to solve the equation (2) for the model, material parameters need to be defined. These material parameters are:

- σ electrical conductivity,
- ϵ_r relative permittivity
- μ_r relative permeability.

These parameters are needed to be assigned to the each of the model domains. The physical meaning of the material properties assignment means, that the equation (2) which describes the electrical component of the electromagnetic field computed in D2 and D3 domains, will be computed in these domains with different parameters. The parameter assignment reflects the physical reality that electromagnetic waves will propagate through air (D2) until it reaches the person (D3).

The domain D2 was assigned with following parameters:

- $\sigma = 8 \times 10^{-15}$ S/m,
- $\epsilon_r = 1$,
- $\mu_r = 1$.

The domain D3 represents the model of the human body which in is general contain many tissues of different organs where each tissue has different electrical properties. The model domain D3 was divided into two parts. The first part is the boundaries of the domain D3 which represents the human skin. For simplicity, the interior of the model domain D3 was assumed to contain primarily muscular tissue. The electrical parameters assigned to the D3 model domain boundaries which represent the skin tissue have been assigned with following parameters:

- $\sigma = 0.8997$ S/m,
- $\epsilon_r = 40.9$,
- $\mu_r = 1$.

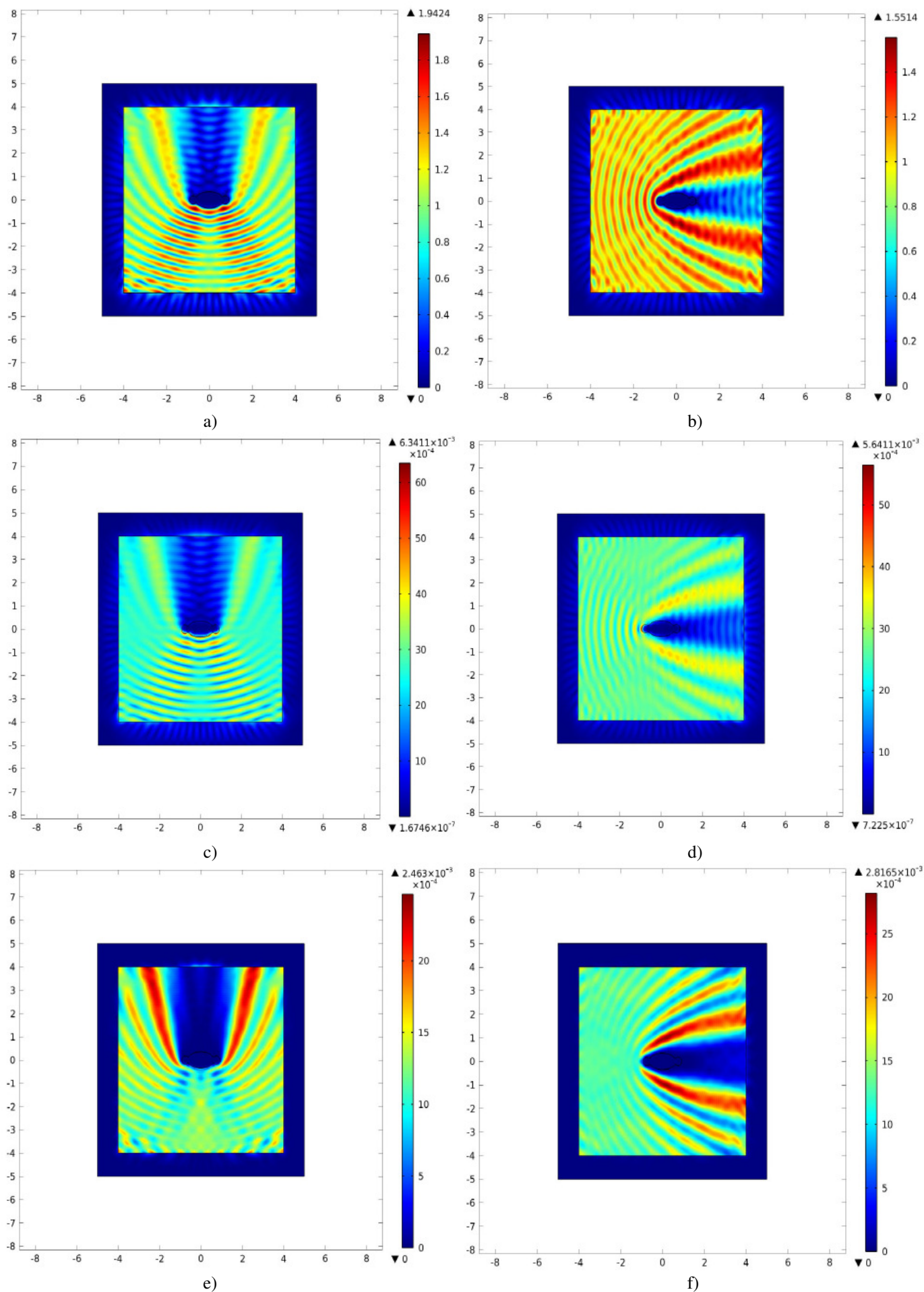


Fig. 2. Experimental results of the shadowing effect simulation: a),b) Intensity of the electric component of the electromagnetic field, c),d) Intensity of the magnetic component of the electromagnetic field, e),f) Energy flux density of the electromagnetic field

The interior of the D3 model domain which represents the muscular tissue have been assigned with these electrical parameters:

- $\sigma = 0.9782 \text{ S/m}$,
- $\varepsilon_r = 54.811$,
- $\mu_r = 1$.

IV. SIMULATION RESULTS

The simulation results show the interaction between the radar emitted electromagnetic field and the model of the human body. The amplitude of the emitted radar signal was 1 V with frequency of 1GHz. The incoming radar transmitted electromagnetic field towards the human being has been analyzed for two different directions.

In these simulation the analysis of the propagation of the electromagnetic fields have shown, that the shadowing effect exists. The person which is located in front of the transmitting radar antenna is clearly creating the zone, where there is high attenuation of radar emitted signals. There is also strong correlation between the angle of incidence of the electromagnetic field and the shape of the shadow zone.

The overall results of the simulation of the shadowing effect are shown in the Fig. 2a) – f). The Fig. a) and Fig. 2 b) display the intensity of electric component E of the electromagnetic field. The intensity of the E component behind the person is greatly reduced by its presence. By comparison of these two results it can be observed that there is the big difference between the directions of incidence of the electromagnetic field with respect to the person. The Fig. 2a) display the electromagnetic field coming towards the person from the bottom while Fig. 2 b) depicts the electromagnetic field coming from aside of the person. The results show, that the shadow zone caused by the person is clearly larger in the first case.

The results portrayed in the Fig 2. c) and Fig. 2 d) show the magnetic component of the electromagnetic field. The drop of the intensity of the magnetic field is similar as in the case of the computed electric field. By closer inspection of these results can be observed, that the intensity of the magnetic field is increased inside the human body. It is caused by the absorption of the electromagnetic field by the skin and muscular tissues of the human body.

The third pair of pictures (Fig. 2 e) and Fig.2 f)) shows the directional energy flux of the overall electromagnetic field. These results show the electromagnetic energy reflected and absorbed by the person. The presence of the person causes the significant drop of electromagnetic energy behind the person.

V. CONCLUSION

The simulation results obtained in the Comsol Multiphysics simulation interface have confirmed that the presence of the human being in the electromagnetic field causes significant drop of energy levels behind it. This means that if a person is situated in front of any of the radar antennas it can create a shadow zone. The shape of dead zones (shadows) strongly depends on the dimensions of the person. The results obtained

by the simulation also show that muscle and skin tissues absorbs the portions of electromagnetic energy which is also contributes to the creation of the shadow zones.

ACKNOWLEDGMENT



We support research activities in Slovakia / Project is co-financed from EU funds. This paper was developed within the Project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055. Also this work was supported by the Slovak Cultural and Educational Grant Agency (KEGA) under the contract No. 010TUKE-4/2012.

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.
- [3] Springer, Jan. 2010, p. 3 18. [3] S. Z. Gurbuz, W. L. Melvin, and D. B. Williams, "Comparison of radarbased human detection techniques," in *Proc. 41st Asilomar Conf. on Signals, Systems and Computers*, 2007, pp. 2199–2203.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] T. Dogaru, L. Nguyen, and C. Le, "Computer Models of the Human Body Signature for Sensing Through the Wall Radar Applications", *Sensors and Electron Devices Directorate, ARL*, September 2007
- [9] FEMLAB 3 –User's Guide, COMSOL AB, January 2004
- [10] <http://www.comsol.com>
- [11] D. Kocur, J. Rovňáková and D. Urdzík, "Short-Range UWB Radar Application: Problem of Mutual Shadowing between Targets" *Elektrorevue*. Vol. 2, no. 4 (2011), pp. 37–43.
- [12] D. Kocur, J. Rovňáková and D. Urdzík, "Mutual shadowing effect of people tracked by the short-range UWB radar", *TSP 2011 : 34th International Conference on Telecommunications and Signal Processing*, August 18–20, 2011 Budapest, Hungary, pp. 302–306.

Study of isotactic polypropylenes by magic angle spinning ^1H nuclear magnetic resonance

¹Peter DURANKA (2nd year), ²Magdaléna UHRÍNOVÁ (2nd year)
Supervisor: ³Dušan OLČÁK

^{1,2,3}Dept. of Physics, FEI TU of Košice, Slovak Republic

¹peter.duranka@tuke.sk, ²magdalena.uhrinova@tuke.sk, ³dusan.olcak@tuke.sk

Abstract— Metallocene isotactic polypropylene (m-iPP) and Ziegler-Natta isotactic polypropylene (ZN-iPP) in pelletized form were studied with the use of the magic angle spinning (MAS) ^1H nuclear magnetic resonance (NMR) spectra. It was found that the MAS contributes to the narrowing of the amorphous lines, which overlap the broad crystalline lines and to the splitting of the central amorphous peak into three sharp peaks at elevated temperatures. The peaks were assigned to the particular proton groups of iPP. The splitting made possible to obtain information on the motion of particular iPP proton groups. Higher mobility of m-iPP chains when compared with that of ZN-iPP chains was deduced from narrower peaks in spectra for m-iPP.

Keywords—isotactic polypropylenes, MAS ^1H NMR spectra, glass transition relaxation, melting process

I. INTRODUCTIONS

Almost all of the newer methods developed for the studies of the structure and molecular dynamics in solids utilize the high spectral resolution provided by magic-angle spinning (MAS) which averages dipole-dipole and anisotropic chemical shift interactions [1, 2].

Properties and microstructure of isotactic polypropylene (iPP) may be directly affected by the polymerization process. One of the methods that affect the properties and microstructure of polypropylene is metallocene catalysis. Metallocene polypropylenes when compared with those prepared by the classical Ziegler-Natta catalysis are much more homogenous as far as their molecular weight, tacticity and distribution of tacticity are concerned and their chains are more alike [3, 4].

The different metallocene iPP (m-iPP) configurational structure reflects in a lower melting temperature when compared with Ziegler-Natta iPP (ZN-iPP), and in the relaxation processes, which were frequently studied by the dynamic mechanical analysis [5, 6, 7].

The aim of this paper is to compare the molecular dynamics of the partially crystalline isotactic polypropylenes prepared by the two different technologies. A comparison was made by means of the MAS ^1H NMR spectra measured within the broad temperature range.

II. THEORETICAL BACKGROUND

For nuclei with spin $1/2$, the key to obtaining a high resolution solid state NMR spectrum depends on the removal of the dipolar interaction for the observed spin. Several ways have been invented for this purpose, each with its own success. The MAS is the simplest scheme and requires only a single pulse while spinning the sample to get the free induction decay (FID). The difficulty is to build a spinning module capable of spinning at around 100 kHz. The central idea of removing the dipolar interaction is based on the fact that it is a traceless anisotropic interaction in nature. Magic-angle spinning has the effect of averaging second-rank terms in the nuclear spin hamiltonian to zero, and so removing the effects of chemical shift anisotropy and dipolar interactions from the NMR spectrum [8].

III. EXPERIMENTAL

A Samples

The samples of predominantly isotactic iPP, METOCENE HM 562 N and TATREN HG 1007 prepared in pelletized form were used for the study. The first sample was prepared by metallocene-catalysed polymerisation and the second sample by polymerisation using classical Ziegler-Natta catalysis. The crystallinities and melting temperatures determined by DSC are of 52 % and 145.2 °C for m-iPP and 55 % and 163.6 °C for ZN-iPP, respectively. The glass transition temperatures determined by DMTA are of 12 °C and 10 °C of m-iPP and ZN-iPP, respectively. More detailed information on the materials under study was reported in [9].

B NMR measurements

Varian solid state NMR spectrometer working at the resonant frequency of 400 MHz was used for the measurement of all MAS ^1H NMR measurements. The temperature range of performed measurements was 30 – 190 °C. For calibration of the temperature inside the rotor the temperature dependence of the ^{207}Pb NMR chemical shift for lead nitrate was used [10]. The experiments were carried out with a probe-head using the 4 mm rotor under the magic angle spinning conditions at the spinning rate of 10 kHz. The chemical shifts were referenced

to the TMS using adamantan as an external standard and they were determined with accuracy better than 0.02 ppm. The duration of the 90° radio-frequency pulse was 2.7 μ s and the recycle delay was 6 s.

IV. RESULTS

The broad-line ^1H NMR spectra of m-iPP recorded at 22°C and 60°C and MAS ^1H NMR spectrum measured at 60°C are depicted in Figure 1. The m-iPP sample consists of amorphous and crystalline regions. Despite this fact, the broad-line ^1H NMR spectrum measured at room temperature shows only one broad line with the width of 27.0 kHz. It follows from this spectrum, that not only crystalline but also amorphous main chains are rigid at room temperature. At 60 °C the chain mobility in amorphous regions increases, which can be observed as line narrowing in the broad-line ^1H NMR spectrum and the ^1H NMR spectrum can be considered as an overlapping of broad and narrow lines related to the crystalline and amorphous chains, respectively. The MAS ^1H NMR spectrum measured at 60 °C unlike the broad-line ^1H NMR spectrum shows much narrower line related to the amorphous chains [8, 11]. The averaging of inter- and intra-molecular interactions is responsible for the narrowing of the central peak in the spectrum measured at 60 °C [11, 12].

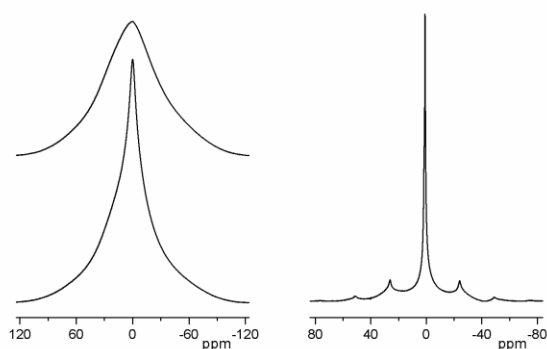


Fig. 1 The broad-line ^1H NMR spectra (left) measured at 22°C (top) and 60°C (bottom) and MAS ^1H NMR spectrum (right) measured at 60°C on m-iPP sample

Fig. 2 shows the spectra measured at three different temperatures. The rise of the temperature results in the chain mobility increase and on that account to the averaging of the dipolar interactions [8]. Due to the rotation of the sample, the central peak splits into three lines with increasing temperature. Particular lines correspond to the hydrogen nuclei in CH_3 , CH_2 and CH groups of iPP. The spectra shown in Fig. 1 indicate that the rotation of the sample under magic angle spinning allows the study of molecular dynamics on the local level. The split lines are superimposed by a very broad crystalline line and this is the reason why the crystalline line cannot be depicted together with the detailed structure of narrow lines in Fig. 2. The chemical shifts of the peaks detected by the solid state MAS ^1H NMR agree with those in the liquid ^1H NMR spectrum of iPP dissolved in 1, 2-dichlorobenzene [4]. The peaks observed at 1.61, 1.30 and 0.91 ppm can be then assigned to methine protons, protons in methylene groups,

which are in anti positions to the adjacent methyl groups, and to the methyl protons and methylene protons in syn positions to the adjacent methyl groups, respectively. The higher temperature the better resolution of the peaks in the spectra and the relative peak intensities 1.00:0.96:3.72 of the completely separated peaks positioned at 1.61, 1.30 and 0.91 ppm, respectively, calculated from the areas of the particular peaks, agree well with the number of protons related to the particular peaks of the MAS ^1H NMR spectra.

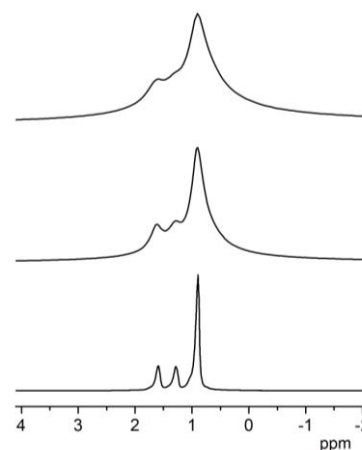


Fig. 2 MAS ^1H NMR spectra of m-iPP measured at 73°C (top), 110°C (middle) and 153 °C (bottom) that is a value above the melting temperature.

The spectra measured for ZN-iPP using MAS ^1H NMR are very similar to those measured for m-iPP in the whole measured temperature range, however, some differences were found.

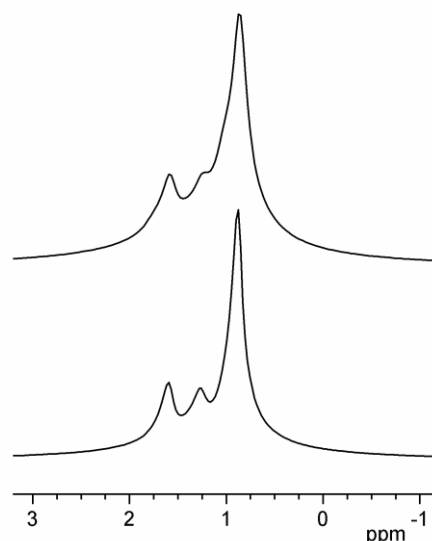


Fig. 3 ^1H NMR spectra for m-iPP (bottom) and ZN-iPP (top) both measured at 100 °C under MAS at spinning frequency of 10 kHz

The MAS ^1H NMR spectra for m-iPP split depending on temperature into two or three lines at elevated temperatures. However, the spectra measured at MAS frequency of 10 kHz for the both samples at the same temperature (Fig. 3) show

that the splitting of the central peak of the spectrum for m-iPP is observed at lower temperature than for ZN-iPP. The differences were also found in the widths of particular peaks. Lines widths of all peaks in m-iPP spectra are narrower than lines widths of peaks from the spectra of ZN-iPP measured at the same temperature (Fig 3.).

V. CONCLUSION

The MAS gives rise to the narrowing of the amorphous line superimposed on a broad crystalline line in the spectra measured at temperatures up to 70 °C. The splitting of the central amorphous peak first into two and then into three sharp peaks was observed at higher temperatures. The peaks at 1.61, 1.30 and 0.91 ppm were assigned to methine protons, protons in methylene groups, which are in *anti* positions to the adjacent methyl groups, and to the methyl protons and methylene protons in *syn* positions to the adjacent methyl groups, respectively. The splitting of the amorphous peak made possible to obtain information on the motion of particular iPP proton group. Higher mobility of m-iPP chains when compared with that of ZN-iPP chains was deduced from narrower lines widths of all peaks in m-iPP spectra when compared with those obtained for ZN-iPP.

REFERENCES

- [1] E.R. Andrew, A. Bradbury, R.G. Eades, Nuclear magnetic resonance spectra from a crystal rotated at high speed, *Nature* 182(1958) 1659
- [2] J. Schaefer, E.O. Stejskal, 13C-NMR of polymers spinning at the magic angle, *J. Am. Chem. Soc.* 98 (1976) 1031–1032.
- [3] Hatakana, T., Mori, H, and Terano, M., *Polymer Degradation and Stability* 64 (1999) 313 – 319
- [4] Busico, V. and R. Cipullo, *Progress in Polymer Science*, 26 (2001) 443 – 533
- [5] Gómez-Elvira, J. M., Tiemblo, P., Elvira, M., Matisova-Rychla, L. and Rychly, J. *Polymer Degradation and Stability*, 85 (2004), 873 – 882
- [6] Hoyos, M., Tiemblo, P. and Gómez-Elvira, J. M., *Polymer* 48 (2007) 183 -194
- [7] Arranz-Andrés, J., Peña, B., Benavente, R., Pérez, E. and Cerrada, M. L., *Eur. Polym. J.*, 43 (2007) 2357 – 2370
- [8] Duer, M. J.: *Solid-State NMR Spectroscopy Principles and Applications*, First published 2002, 0-632-05351-8
- [9] Fričová, O., Uhrínová, M., Hronský V., Kovaľáková, M., Olčák, D., Chodák, I., Spěvácěk, J., *Express. Polym. Lett.*, 6 (2012), 204 – 212
- [10] Bielecki, A., Burum, D. P., *J. Magn. Reson. Ser. A.*, 116 (1995) 215 – 220
- [11] Duer, M. J.: *Introduction to Solid-State NMR Spectroscopy*, Blackwell Publishing, 2004, p. 62, 293
- [12] Kitamaru R., *Nuclear Magnetic Resonance Principles and Theory*, Elsevier, 1990, 122, 213
- [13] Hedesiu, C., Demco, D. E., Kleppinger, R., Vanden Poel, G., Gijsberg, W., Blümich, B., Remerie, K., Litvinov, V. M., *Macromolecules* 40 (2007) 3977 – 3989
- [14] Geppi, M., Harris, R.K., Kenwright A.M. and Say, B.J., *Solid State Nuclear Magnetic Resonance*, 12 (1998), 15 – 20

Section:
Electrical & Electronics Engineering
Poster form

A Novel Approach to Dynamic Emulation of Mechanical Loads Based on Industrial Converters' Features

¹Karol KYSLAN (3rd year), ²Matúš HRIC (3rd year), Supervisor: ³František ĎUROVSKÝ

^{1,2,3} Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

¹karol.kyslan@tuke.sk, ²matus.hric@tuke.sk, ³frantisek.durovsky@tuke.sk

Abstract—The paper presents and analyses a novel method for the dynamic emulation of mechanical loads. This approach is based on industrial converters' features and can be applied in design, testing and validation of the speed controllers for complex mechatronic systems. Mechanical system as a prototype which has to be controlled is replaced by the torque controlled electrical drive acting as active load. The reference torque for the active load is calculated in the closed loop algorithm. The experimental results of linear load are presented.

Keywords—dynamic emulation, industrial converter, RT-Lab, test rig

I. INTRODUCTION

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, output data of imitated and imitating system have to be exactly the same. The imitating system is called emulator.

Presented dynamic emulation of mechanical loads is the emulation technique which can be used for design and validation of control algorithms for speed and torque drives in mechatronic applications. The experimental setup of presented

approach is shown in Fig. 1. For the drive under test the speed control algorithm has to be found. This drive will be in factory conditions loaded by the mechanical load. In the laboratory facility, this mechanical load is replaced by the load machine (dynamometer), which emulates the behaviour of the real mechanical load. Load drive and drive under test are coupled by the rigid shaft. The parameters of the mechanical load are represented by a mathematical model built in MATLAB Simulink environment in Command Station, and running in real-time after compiling to C-code and loading to Target-PC. The meaning of the signals is described in Chapter 2. Opal-RT software RT-Lab based on the QNX Linux OS is used for the real-time control. CAN-ACx data acquisition card is used for CAN communication between converter and the personal computer. Both converters are able to work in 4-quadrant operation conditions and in addition both converters are equipped with SLB board for fast communication with SiMoLink (Siemens Motion Link) protocol. Moreover, DC master converter is equipped with CBC board (Communication Board CAN) for communication with CAN protocol. This arrangement enables the real-time control of both industrial converters and creates the test rig, which is being used for various experiments and measurements.

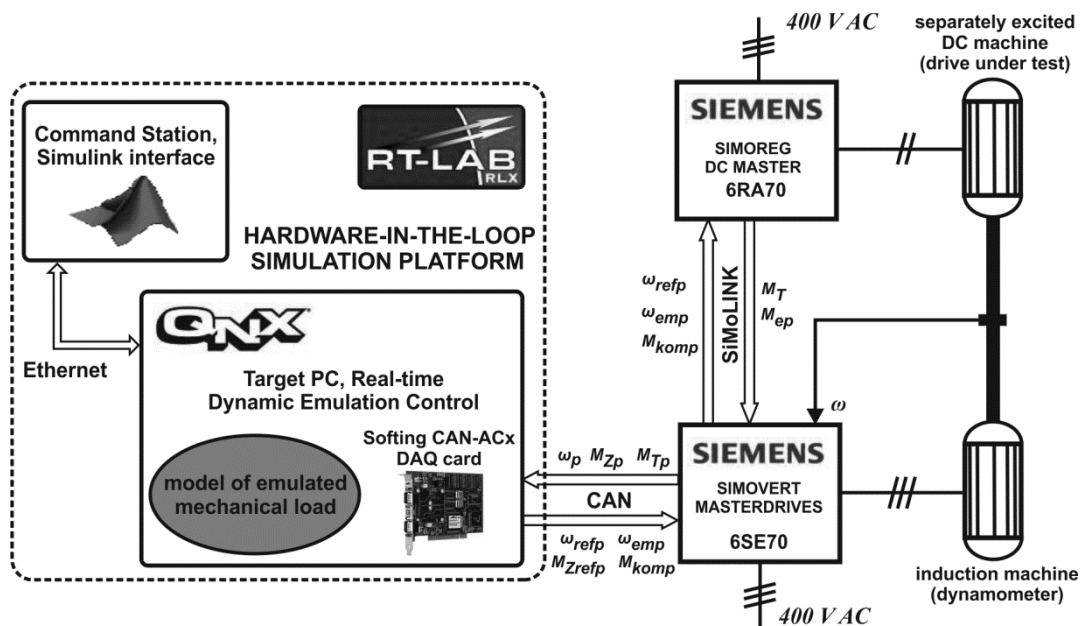


Fig. 1. Experimental setup

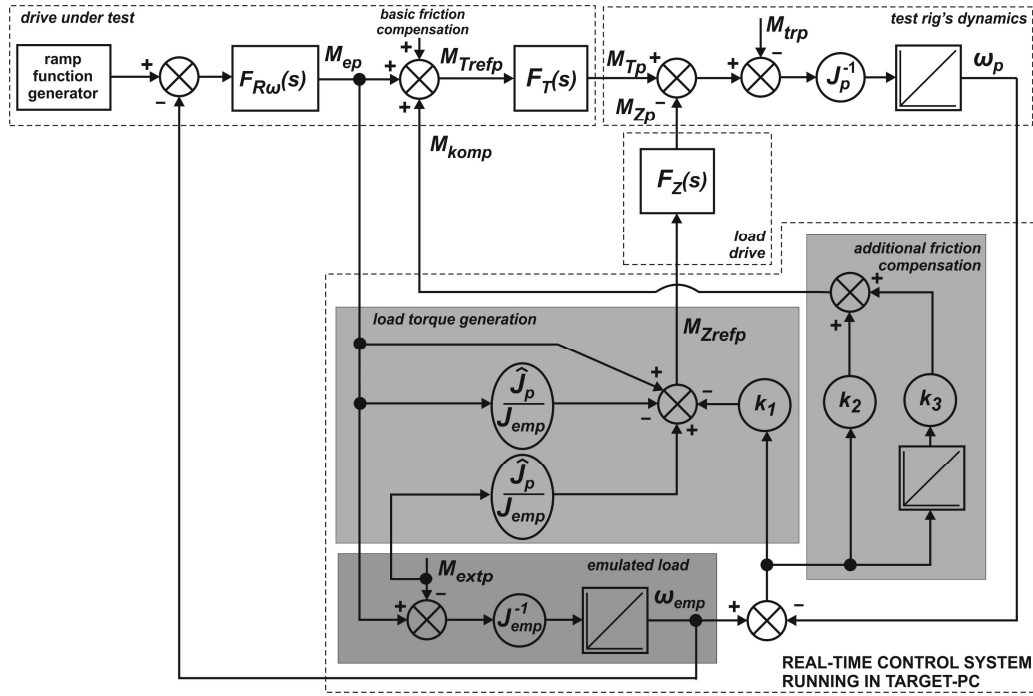


Fig. 2. Control algorithm scheme for proposed method

All the measurements and computations were executed in 250μs, which is satisfactory sampling time for current and speed control loops. Current methods used for dynamic emulation of mechanical loads can be found in [4], [5] and [6]. But proposed method uses a novel control structure utilizing the available features of modern industrial converters.

II. PRESENTED APPROACH

Drive control of the most of industrial drive converters is performed in p.u. formulations and it is the same for whole power range of drives from kW to MW. Thus control algorithm was formulated in p.u. variables according to these equations:

$$M_{Tp} = M_T \frac{1}{NM_T}; \quad M_{Zp} = M_Z \frac{1}{NM_Z}; \quad (1)$$

M_{Tp} - normalized torque of drive under test [-],

M_T - drive under test absolute torque [Nm],

NM_T - the norm of the drive under test [Nm],

M_{Zp} - normalized torque of load drive [-],

M_Z - load drive absolute torque [Nm],

NM_Z - the norm of the load drive [Nm],

$$\omega_p = \omega \frac{1}{N\omega}; \quad \omega_{emp} = \omega_{em} \frac{1}{N\omega_{em}}; \quad (2)$$

ω_p - normalized speed of drive under test [-],

ω - drive under test absolute speed [s^{-1}],

$N\omega$ - the norm of the drive under test speed [s^{-1}],

ω_{emp} - normalized speed of the emulated load [-],

ω_{em} - absolute speed of the emulated load [s^{-1}],

$N\omega_{em}$ - the norm of the emulated load speed [s^{-1}].

Control algorithm is shown in Fig. 2. Per-normal value of moment of inertia was formulated according to [1]. When it is

used in the control structure, it enables to emulate inertias with much higher values, than the value of the test rig inertia:

$$J_p = J \frac{N\omega}{NM_T}; \quad \left[s; kg.m^2; \frac{s^{-1}}{Nm} \right] \quad (3)$$

J - nominal value of the total test rig inertia (i.e. drive under test, load drive and clutch)

J_p - per-normal value of the total test rig inertia.

Per-normal inertia is the time in seconds that it takes to accelerate the motor and load inertia from zero speed to motor base speed with motor rated torque and the given reflected inertia [1]. The full control law for the reference torque of load drive is stated as:

$$M_{Zrefp} = M_{Trefp} - M_{Trefp} \frac{\hat{J}_p}{J_{emp}} + M_{extp} \frac{\hat{J}_p}{J_{emp}} + k_1(\omega_{emp} - \omega_p) \quad (4)$$

The derivation of (4) can be found in [3]. Value \hat{J}_p is the estimated of per-normal value of total test rig inertia, obtained by measurement, k_1 is the gain of the P controller used to achieve required model tracking. The real torque loops are in Fig. 2 modelled by first order lags $F_T(s)$ and $F_Z(s)$. $F_{Rw}(s)$ is the transfer function of speed controller, which parameters are tested by different emulated loads. M_{trp} is the input presenting all the linear and nonlinear friction torques in the test rig mechanics.

In Chapter 3, dynamic analysis for linear system behaviour is done. However, the system is non-linear. But linear analysis can be done, if these conditions are fulfilled:

- speed controller do not reach the saturation,
- nonlinear friction torques in the test rig are assumed to be ideally compensated,
- the transport delay between reference and actual values of the control structure and backlash in the shaft connection are omitted.

For the compensation of nonlinear friction torques, “friction compensation” feature of the SIMOREG converter is used [2] and moreover, additional control loop with parameters k_2 and k_3 is introduced into the control structure. In the Fig. 3 it is shown, how this additional control loop works and rejects the friction torques.

III. DYNAMIC ANALYSIS

The operation can be analyzed using the scheme in Fig. 2. In this paragraph both drive under test and load drive torque loops are considered to be unity gain. Main goal is to design and test different speed controllers, and thus these inputs are taken into the consideration: M_{ep} as the output of speed controller, M_{trp} as nonlinear friction input and M_{extp} as desired value of load torque. If the next equation is valid:

$$\omega_{emp}(s) = \frac{1}{sJ_{emp}} [M_{ep}(s) - M_{extp}(s)] \quad (5)$$

transfer function of the speed ω_p is presented as:

$$\omega_p(s) = \frac{\hat{J}_p s^2 + (k_1 + k_2)s + k_3}{J_p s^2 + (k_1 + k_2)s + k_3} \omega_{emp}(s) - \frac{s}{J_p s^2 + (k_1 + k_2)s + k_3} M_{trp}(s) \quad (6)$$

If the value of total inertia is known exactly :

$$\hat{J}_p = J_p \quad (7)$$

(6) is simplified to:

$$\omega_p(s) = \omega_{emp}(s) - \frac{s}{J_p s^2 + (k_1 + k_2)s + k_3} M_{trp}(s) \quad (8)$$

From (5) and (8) it is obvious, that if (7) is fulfilled the actual speed of drive under test is equal to the emulated speed, and our goal is reached. However, real speed of drive under test is affected by presence of linear and nonlinear torques M_{trp} . Its compensation is done in converter and additional compensation is done by control algorithm with suitable choice of controller's parameters. Controller's parameters in Fig. 4 were chosen so, that even if friction compensation in converter was unsatisfactory, influence of the unwanted friction would be rejected by dynamic emulation control with the gain of at least -25dB.

Characteristic equation of the system, based on (14), is:

$$s^2 + \frac{k_1 + k_2}{J_p} s + \frac{k_3}{J_p} = 0 \quad (9)$$

If (9) is compared to the general equation of the 2nd order element:

$$s^2 + 2d\omega_0 s + \omega_0^2 = 0 \quad (10)$$

where d presents the damping and ω_0 the natural frequency of the 2nd order element, we get:

$$\frac{k_1 + k_2}{J_p} = 2d\omega_0; \quad \frac{k_3}{J_p} = \omega_0^2 \quad (11)$$

Thus values of the controller parameters can be calculated as:

$$\begin{aligned} k_1 &= 2d\omega_0 J_p - k_2 \\ k_2 &= 2d\omega_0 J_p - k_1 \\ k_3 &= J_p \omega_0^2 \end{aligned} \quad (12)$$

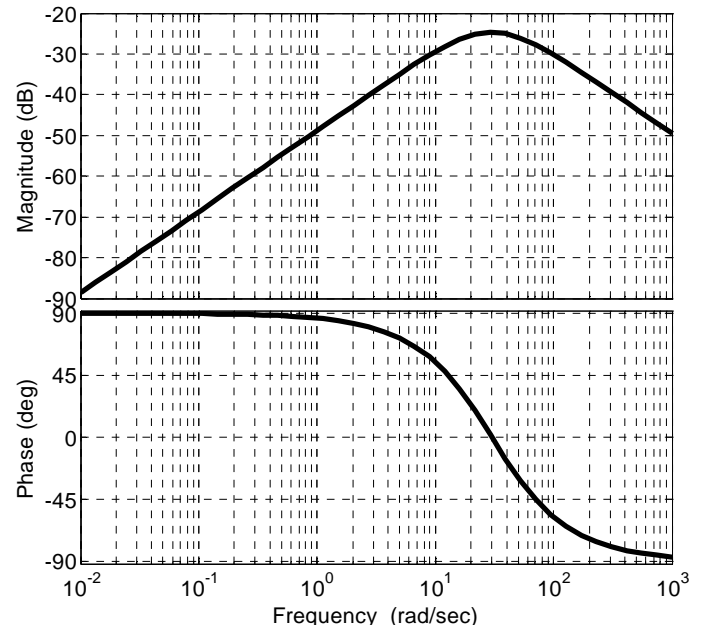


Fig. 3. Bode diagram of ω_p/M_{trp} , friction torques rejection, $d=0,95$; $\omega_0=30 \text{ rad.s}^{-1}$

IV. EXPERIMENTAL RESULTS

In the following chapter the emulation of linear load will be presented. The dynamics of the most simple linear load is presented in (5). Per-normal inertia of emulated load was increased to the value:

$$J_{emp} = 3J_p \quad (13)$$

and no additional torque was applied:

$$M_{extp} = 0 \text{ Nm} \quad (14)$$

Results are presented in Fig. 4. In order to avoid nonlinear effects the ramp function for the reference speed ω_{refp} was chosen in such a way that the speed controller output does not reach the saturation limit. Behaviour of load torque is expected. Until the time $t=8\text{s}$ load torque equals zero. In the time instant $t=8\text{s}$, the inertia was increased according to (13) and speed controller remain unchanged. It can be observed, that without real mechanical load connected to the drive under test, the higher torque value is required and the overshoot of speed is increased.

Unfortunately, the backlash effect can be observed in Fig. 4. on the real torque and speed values. It is present because of not ideally rigid connection of motor shafts. We can avoid this unwanted effects by either using a better test rig mechanics or by implementing backlash rejection control into the control structure.

Speed error shown in the middle part of Fig. 4 always stays within the band of $\pm 2\%$ of nominal speed value. This quality of emulation is comparable with the results obtained in [4] – [6].

V. CONCLUSION

Novel speed control structure for dynamic emulation of mechanical loads is presented in this work. Using of per-normal inertia calculus enables emulating of inertia with much higher values than the value of the test rig inertia. In Chapter 4 the dynamic analysis has been done and it's been

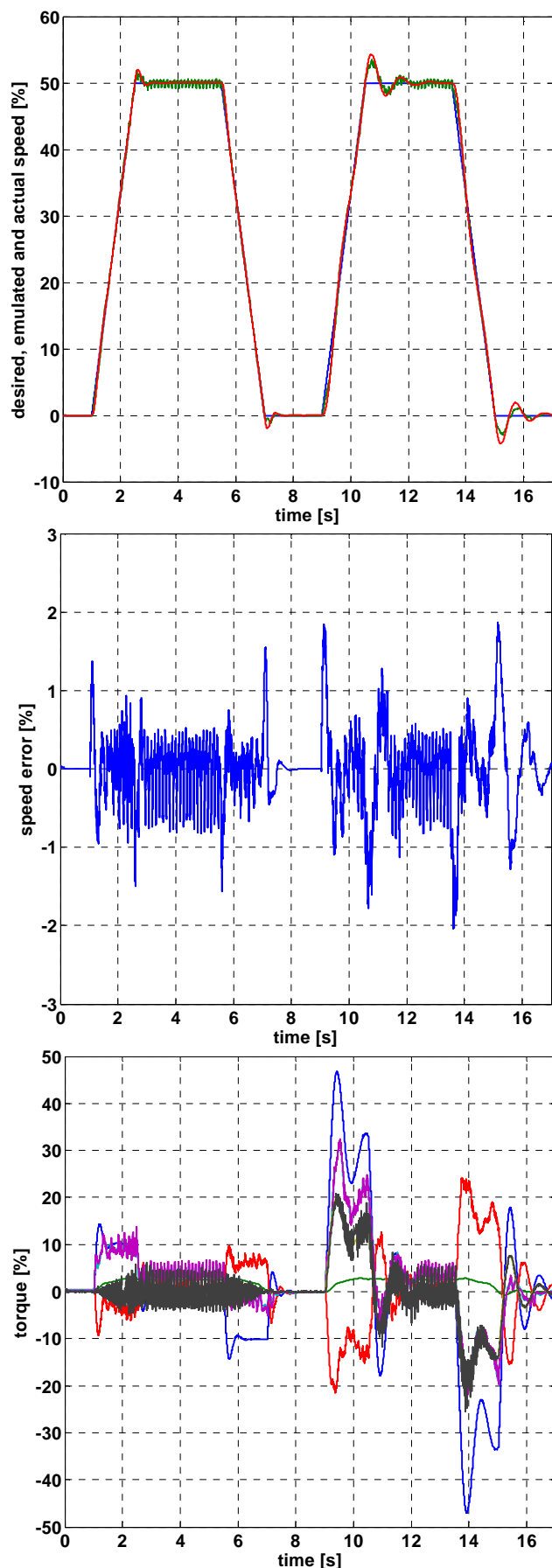


Fig. 4. Emulation of linear load, $J_{emp} = 3J_p$

top: desired speed (blue), ω_p (green), ω_{em} (red);
middle: speed error ($\omega_{em} - \omega_p$);
bottom: M_{ep} (blue), M_{komp} (red), M_T (purple), M_L (grey),
 basic friction compensation (green).

shown the method preserves pole-zero structure of emulated load. The performance of the method was tested with the emulation of linear load and very good speed tracking was obtained. Method is based on the industrial converters' features, but it can be used with any other type of converter, with implemented friction compensation.

The research on this kind of the test rig control is important for various measurements and experiments, at our department mainly focused on the testing of a cycloid gearboxes [7]. Main advantage of proposed approach is that there is no need to have various mechanical loads in the laboratory, because their behaviour is emulated by the torque-controlled drive, what is fast, user-friendly and effective solution.

ACKNOWLEDGMENT

This work is the result of the project implementation: *Výskum modulov pre inteligentné robotické systémy*, ITMS: 26220220141, supported by the Research & Development Operational Programme funded by the ERDF. This work was also supported by the Slovak Research and Development Agency under the contract No. APVV-0185-10.



REFERENCES

- [1] Boulter, B.T.; "Applying drive performance specifications to systems applications. I.Speed performance," *Industry Applications, IEEE Transactions on*, vol.37, no.4, pp.1082-1087, Jul/Aug 2001.
- [2] SIEMENS Simoreg DC Master compendium, available online: <http://support.automation.siemens.com/WW/llisapi.dll?func=cslib.csinf&o&lang=en&objID=10804967&subtype=133300> (12/ 2011).
- [3] Kyslan, K.; Ďurovský, F.; „An Approach to Dynamic Emulation of Mechanical Loads“ In: *17th International Conference on Electrical Drives and Power Electronics, EDPE 2011*. 28. - 30.9.2011, Stará Lesná, Slovakia. 2011. s. 308-312. ISBN 978-80-553-0734-3.
- [4] 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.
- [5] Žalman, M.; Macko, R.; „Design and realization of programmable emulator of mechanical loads," *16th IFAC World Congress, Czech Republic*, vol. 16, 2005. pp. 246-250.
- [6] Rodic, M.; Jezernik, K.; Trlep, M.; "Mechatronic Systems' Control Design Using Dynamic emulation of Mechanical Loads," *Automatika: Journal for Control, Measurement, Electronics, Computing and Communications*, vol. 47, no 1-2, pp. 11-18, ISSN 0005-1144, May 2006.
- [7] Hric, M.; Fedák, V.; Ďurovský, F.; „Properties Analysis of Precise Gearboxes for Robotic Applications," *In: 17th International Conference on Electrical Drives and Power Electronics, EDPE 2011: 5th Joint Slovak-Croatian Conference*: 28. - 30.9.2011, Stará Lesná, Slovakia. 2011. s. 318-322. - ISBN 978-80-553-0734-3.

A Novel Multiphase Boost Converter with High Efficiency of Energy Conversion

¹Ján PERDULAK (2nd year), ²Oksana KHREBTOVA, ³Oleksii LESHCHUK
Supervisor : ⁴Dobroslav KOVÁČ

^{1,4}Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic
^{2,3}Institute of Electromechanics, Energy Saving and Control Systems, Kremenichuk Mykhailo Ostrohradskyi National University, Ukraine

¹jan.perdulak@tuke.sk, ⁴dobroslav.kovac@tuke.sk

Abstract — This article introduce a novel concept of boost converter with high efficiency of energy conversion. This new concept allows effective utilization of energy from photovoltaic solar cell. The effective utilization of energy is ensured by adding five parallel legs to the conventional boost converter with one leg. The simulation model has been built and the simulation results obtained to verify the theoretical properties of multiphase boost converter.

Keywords — multiphase boost converter, photovoltaic, SLPS interface, energy conversion, CCM mode

I. INTRODUCTION

This paper presents the novel concept of multiphase boost converter with high efficiency of energy conversion. The high efficiency of energy conversion is ensured by adding five more parallel legs to the conventional boost converter with one leg. The suitable algorithm of switches control in particular legs ensures that the almost whole PV output energy from the PV panel is effective utilized.

II. EFFICIENCY OF ENERGY CONVERSION

Fig.1. explains the problem of efficiency of energy conversion. The impinging sun energy P_{INsun} is converted by PV module direct to the electric energy. According to the material which PV module is build this conversion efficiency moving from 5% (a-Si) to 30% (GaAs), [1]. The output PV energy P_{OUT_PV} equals the input energy to the converter P_{IN_con} .

$$P_{OUT_PV} = P_{IN_con} \quad (1)$$

Only a part of this input energy P_{IN_con} is drawn by the converter system. The converter works in switching mode with any set value of duty cycle z . According to the set value of duty cycle z , the real amount of input energy to the converter is

$$P_{IN_con}^* = P_{IN_con} \cdot z \quad (2)$$

It can be seen that the real amount of input energy to the converter $P_{IN_con}^*$ is less that the P_{IN_con} because the duty cycle z is theoretically moving from 0 to 1.

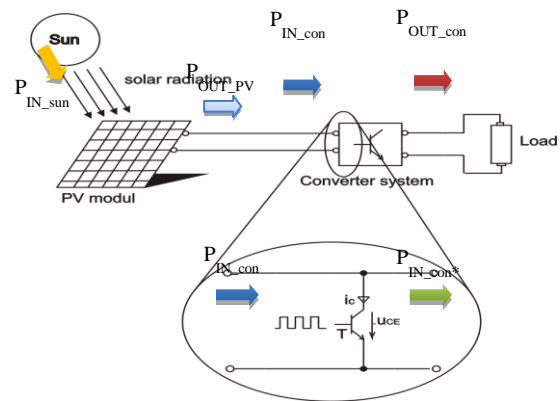


Fig. 1. Overview at efficiency of energy conversion.

The three different efficiencies we can define as:

The conversion efficiency of PV module

$$\eta_{PV} = \frac{P_{OUT_PV}}{P_{IN_sun}} \quad (3)$$

The converter efficiency

$$\eta_{con} = \frac{P_{OUT_con}}{P_{IN_con}^*} \quad (4)$$

The efficiency of energy conversion

$$\eta_E = \frac{P_{OUT_con}}{P_{OUT_PV}} \quad (5)$$

where P_{OUT_con} is the output converter energy.

Nowadays, the efficiency of the soft switching DC/DC converters η_{con} is very well. It is moving around the 97%. But on the other hand the efficiency of energy conversion η_E is in comparison with converter efficiency η_{con} much lower. This fact belongs between one of major factor of long-term energy recovery and high cost of PV modules. One way to reduce the long-term energy recovery and so high cost of PV cells is proposed multiphase boost converter which ensures the

equality of converter efficiency η_{con} and the efficiency of energy conversion η_E .

III. THE PROPOSED CONCEPT OF MULTIPHASE BOOST CONVERTER

The proposed topology of multiphase boost converter is in fig.2. The multiphase boost converter has, in comparison with the conventional boost converter with one leg, five more parallel legs with five inductors ($L_2 - L_6$), five rectifier diodes ($D_{21} - D_{61}$) and five switches ($S_{21} - S_{61}$). There are also six auxiliary switches $S_{12} - S_{62}$ on the converter input. The auxiliary switches $S_{12} - S_{62}$ ensure the connection of the input voltage U_{IN} to the load Z . The topology of multiphase converter is also complemented by six auxiliary diodes $D_{12} - D_{62}$ which serve as freewheeling diodes. The freewheeling diodes $D_{12} - D_{62}$ return the inductor storage energy $W_{L1} - W_{L6}$ back to the load Z after the particular complementarily switches $S_{a1} - S_{a2}$ are turned off (where $a \in \{1 - 6\}$).

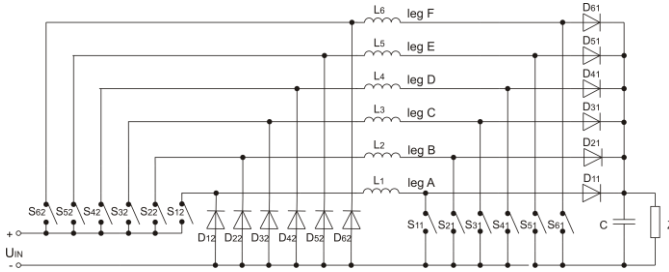


Fig. 2. Multiphase boost converter.

IV. PRINCIPLE OF OPERATION

The proposed multiphase boost converter has 6 operating cycles within each period. The corresponding operation waveforms are shown in fig.3.

Mode 1 ($t_0 - t_1$): The switches S_{11} and S_{12} (leg A) are turned on at the time t_0 . The energy in form of magnetic field begins to accumulate in inductor L_1 . The diode D_{12} is reverse biased so the whole input current is closed in loop $+U_{IN} - S_{12} - L_1 - S_{11} - U_{IN}$. In this mode the switches S_{51} , S_{52} (leg E) and S_{61} , S_{62} (leg F) are in on-state. The input energy is delivering to the inductor L_5 and L_6 in particular legs, too. The switches S_{21} , S_{22} (leg B) and S_{31} , S_{32} (leg C) are in off-state. The inductor energy W_{L2} and W_{L3} is delivered through diodes D_{21} and D_{31} to the load Z . The equivalent equations are:

The inductor voltages $u_{L1}(t)$, $u_{L5}(t)$ and $u_{L6}(t)$ are

$$u_{L1}(t) = u_{L5}(t) = u_{L6}(t) = U_{IN} = L_{1(5,6)} \frac{di_{L1(5,6)}(t)}{dt} \quad (6)$$

The inductor voltage $u_{L2}(t)$, $u_{L3}(t)$ and $u_{L4}(t)$ are

$$u_{L2}(t) = u_{L3}(t) = u_{L4}(t) = -U_{OUT} = L_{2(3,4)} \frac{di_{L2(3,4)}(t)}{dt} \quad (7)$$

The currents flow through inductors L_1 , L_5 , L_6 and co-operating switches $S_{11} - S_{12}$, $S_{51} - S_{52}$, $S_{61} - S_{62}$ are

$$\begin{aligned} i_{S11}(t) &= i_{S12}(t) = i_{S51}(t) = i_{S52}(t) = i_{S61}(t) = \\ i_{S62}(t) &= i_{L1}(t) = i_{L5}(t) = i_{L6}(t) = \\ &= \frac{1}{L_{1(5,6)}} \int_{t_0}^{t_1} u_{L1(5,6)}(t) dt + I_{L1(5,6)}(t_0) = \\ &= \frac{U_{IN}}{L_{1(5,6)}} (t_1 - t_0) + I_{L1(5,6)}(t_0) \end{aligned} \quad (8)$$

The currents flow through inductors L_2 , L_3 , L_4 and couple of diodes $D_{21} - D_{22}$, $D_{31} - D_{32}$, $D_{41} - D_{42}$ are

$$\begin{aligned} i_{D21}(t) &= i_{D22}(t) = i_{D31}(t) = i_{D32}(t) = i_{D41}(t) = \\ i_{D42}(t) &= i_{L2}(t) = i_{L3}(t) = i_{L4}(t) = \\ &= \frac{1}{L_{2(3,4)}} \int_{t_0}^{t_1} u_{L2(3,4)}(t) dt + I_{L2(3,4)}(t_0) = \\ &= -\frac{U_{OUT}}{L_{2(3,4)}} (t_1 - t_0) + I_{L2(3,4)}(t_0) \end{aligned} \quad (9)$$

The inductor current $i_{L1}(t)$ exponentially increase from initial value I_{L1} to the maximum value I_{L1max} (reached at the time t_3) with time constant $\tau_I = L_1/R$.

Mode 2 ($t_1 - t_2$) and **mode 3 ($t_2 - t_3$)** are the same as **mode 1**. Only another co-operating switches $S_{21} - S_{22}$ (leg B, **mode 2**) and $S_{31} - S_{32}$ (leg C, **mode 3**) are turned on, on-state S_{11} , S_{12} , S_{61} , S_{62} (leg A and leg F, **mode 2**) and $S_{11} - S_{12}$, $S_{21} - S_{22}$ (legs A and B, **mode 3**) and off-state $S_{41} - S_{42}$, $S_{51} - S_{52}$ (legs D and E, **mode 3**) and S_{31} , S_{32} , S_{41} , S_{42} (leg D and leg C, **mode 2**). The corresponding equations are the same. Only subscript are changed.

Mode 4 ($t_3 - t_4$): The switches S_{11} and S_{12} are turned off and S_{41} and S_{42} are turned on the beginning of this mode at the time t_3 . The inductor energy W_{L1} begins to deliver through diode D_{11} to the load Z . The polarity of inductor voltage $u_{L1}(t)$ is reversed so the diode D_{12} is in on-state. The output current $i_Z(t)$ is enclosed in the loop $L_1 - D_{11} - Z - D_{12}$. The switches S_{21} , S_{22} (leg B) and S_{31} , S_{32} (leg C) are on-state and the input energy is delivering to the inductor L_2 and L_3 . The switches S_{51} , S_{52} (leg E) and S_{61} , S_{62} (leg F) are in off-state. The inductor energy W_{L5} and W_{L6} is delivered through diodes D_{51} and D_{61} to the load Z . The equivalent equations are:

The inductor voltages $u_{L4}(t)$, $u_{L2}(t)$ and $u_{L3}(t)$ are

$$u_{L4}(t) = u_{L2}(t) = u_{L3}(t) = U_{IN} = L_{4(2,3)} \frac{di_{L4(2,3)}(t)}{dt} \quad (10)$$

The inductor voltages $u_{L1}(t)$, $u_{L5}(t)$ and $u_{L6}(t)$ are

$$u_{L1}(t) = u_{L5}(t) = u_{L6}(t) = -U_{OUT} = L_{1(5,6)} \frac{di_{L1(5,6)}(t)}{dt} \quad (11)$$

The currents flow through inductors L_4, L_2, L_3 and co-operating switches $S_{41} - S_{42}, S_{21} - S_{22}, S_{31} - S_{32}$ are

$$\begin{aligned} i_{S41}(t) &= i_{S42}(t) = i_{S21}(t) = i_{S22}(t) = i_{S31}(t) = \\ i_{S32}(t) &= i_{L4}(t) = i_{L2}(t) = i_{L3}(t) = \\ \frac{1}{L_{4(2,3)}} \int_{t_3}^{t_4} u_{L4(2,3)}(t) dt + I_{L4(2,3)}(t_3) &= \\ \frac{U_{IN}}{L_{4(2,3)}} \left(t_4 - t_3 \right) + I_{L4(2,3)}(t_3) \end{aligned} \quad (12)$$

The currents flow through inductors L_4, L_5, L_6 and couple of diodes $D_{11} - D_{12}, D_{51} - D_{52}, D_{61} - D_{62}$ are

$$\begin{aligned} i_{D11}(t) &= i_{D12}(t) = i_{D51}(t) = i_{D52}(t) = i_{D61}(t) = \\ i_{D62}(t) &= i_{L1}(t) = i_{L5}(t) = i_{L6}(t) = \\ \frac{1}{L_{1(5,6)}} \int_{t_3}^{t_4} u_{L1(5,6)}(t) dt + I_{L1(5,6)}(t_3) &= \\ -\frac{U_{OUT}}{L_{1(5,6)}} \left(t_4 - t_3 \right) + I_{L1(5,6)}(t_3) \end{aligned} \quad (13)$$

Mode 5 ($t_4 - t_5$) and mode 6 ($t_5 - t_6$) are the same as mode 4. Only another co-operate switches $S_{61} - S_{62}$ (leg F, mode 6) and $S_{51} - S_{52}$ (leg E, mode 5) are turned on, on-state $S_{31}, S_{32}, S_{41}, S_{42}$ (leg C and leg D, mode 5) and $S_{41} - S_{42}, S_{51} - S_{52}$ (legs D and E, mode 6) and off-state $S_{11} - S_{12}, S_{21} - S_{22}$ (legs A and B, mode 6) and $S_{11}, S_{12}, S_{61}, S_{62}$ (leg A and leg F, mode 5). The corresponding equations are the same. Only subscript are changed.

If we assume that the average value inductor voltage $U_{L(AV)}$ has to be zero for period T , equation (14), then the average value of the output voltage $U_{OUT(AV)}$ of proposed topology of boost converter in CCM can be easily derived.

$$U_{L(AV)} = \frac{1}{T} \int_0^T u_L(t) dt = 0 \quad (14)$$

The average value of the output voltage $U_{OUT(AV)}$ of proposed topology of multiphase boost converter in CCM.

$$U_{OUT(AV)} = \frac{z}{1-z} U_{IN} \quad (15)$$

It is clear that the minimum setting of value for duty cycle z has to be 0,5 respectively 50% of period T . If the value of duty cycle z is less than the 0,5 than the average value of the output voltage $U_{OUT(AV)}$ will be smaller than the input voltage U_{IN} . The function of multiphase boost converter will be incorrect in this case.

V. SIMULATION RESULTS

The simulation model of multiphase boost converter shown in fig.4 was created in simulation environment OrCAD Capture CSI to verify its theoretical properties.

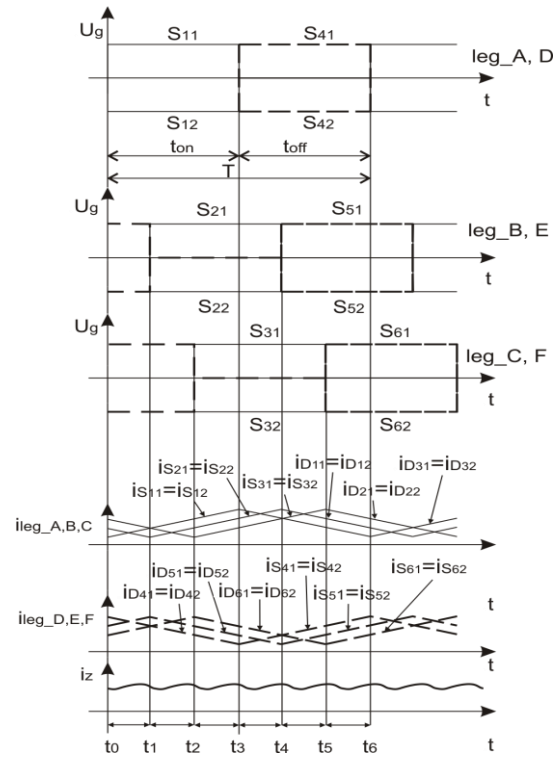


Fig. 3. Theoretical waveforms of proposed multiphase boost converter.

Parameters:

Switching frequency	$f_s = 50 \text{ kHz}$,
output voltage	$U_{bat} = 14 \text{ V}$,
input voltage	$U_{PV} = 10 \text{ V}$,
inductance	$L_1 = L_2 = L_3 = L_4 = L_5 = L_6 = 50 \mu\text{H}$,
capacitance	$C = 22 \mu\text{F}$
duty cycle	$z = 0,6$.

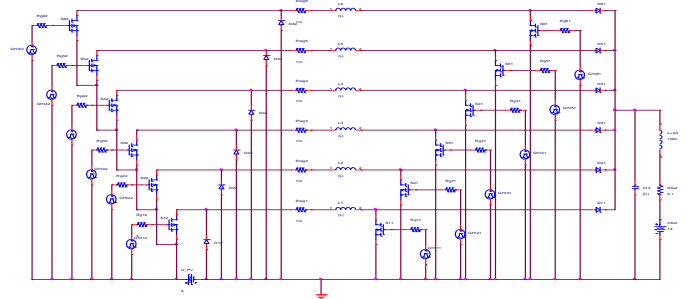


Fig.4. Simulation model of proposed multiphase boost converter.

Fig.5. shows the inductor currents $i_{L1}(t) - i_{L6}(t)$, diode currents $i_{D11}(t), i_{D12}(t) - i_{D61}(t), i_{D62}(t)$ and transistor currents $i_{S11}(t), i_{S12}(t) - i_{S61}(t), i_{S62}(t)$ in particular legs. It can be seen that after turning on co-operating transistors in given leg the inductor begins to accumulate energy in form of magnetic field. This energy is consequently delivered to the load Z till these couples of transistors are turned-off.

Control structure created in Simulink environment is shown in fig.6. On this purpose the PSpice SLPS (SimuLink PSpice) simulation environment, supports the substitution of

an actual Simulink block with an equivalent analog PSpice electrical circuit, was used.

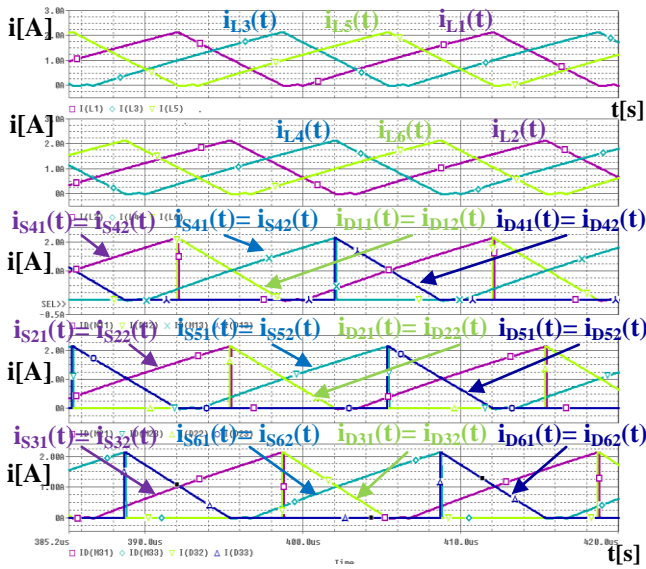


Fig.5. Waveforms of inductor currents $i_{L1}(t)$ till $i_{L6}(t)$, diode currents $i_{D11}(t)$, $i_{D12}(t)$ till $i_{D61}(t)$, $i_{D62}(t)$ and transistor currents $i_{S11}(t)$, $i_{S12}(t)$ till $i_{S11}(t)$, $i_{S12}(t)$.

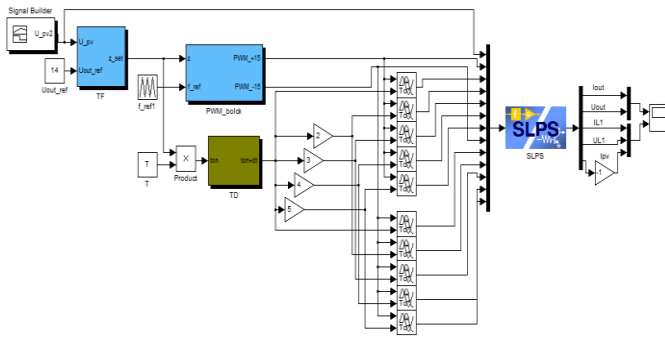


Fig.6. Simulink model of control structure of six phase boost converter.

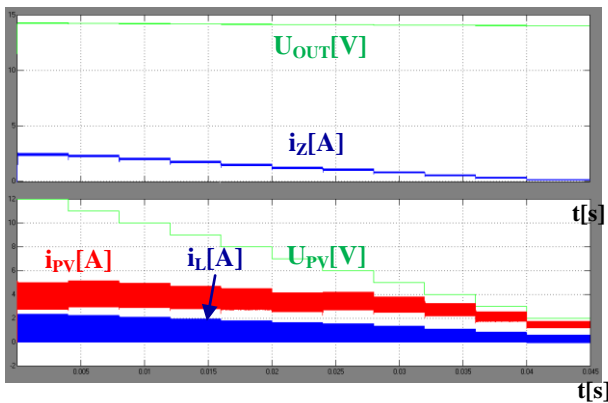


Fig.7. Waveforms of load current $i_z(t)$, output voltage U_{OUT} (upper part), inductor and PV currents $i_{pv}(t)$ and $i_L(t)$ and PV voltage (lower part).

Fig.7. shows overview of load current $i_z(t)$, inductor current $i_L(t)$, input photovoltaic current $i_{pv}(t)$, input and output voltage $u_{out}(t)$ at different values of photovoltaic voltage U_{PV} . The PV voltage is moving in range from 12V to 2V, with decrement 1V. In comparison with conventional boost converter with one leg the load current $i_z(t)$ works in CCM for

whole range of input voltages U_{PV} including the minimum value $U_{PV} = 2V$, Fig.8.

The extended waveforms of inductor current $i_L(t)$, load current $i_z(t)$ and photovoltaic current $i_{pv}(t)$ are shown in fig.9. The peak-to-peak values of load ripple current Δi_z are very small for whole range of input PV voltage U_{PV} . The peak-to-peak values of load ripple current Δi_z is moving around = 150 mA .

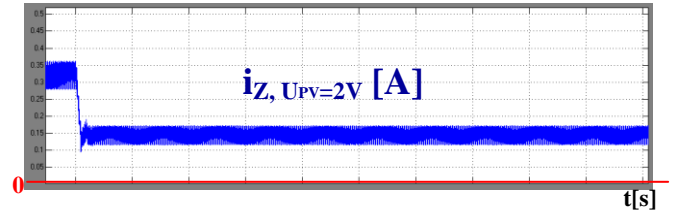


Fig.8. Load current $i_z(t)$.

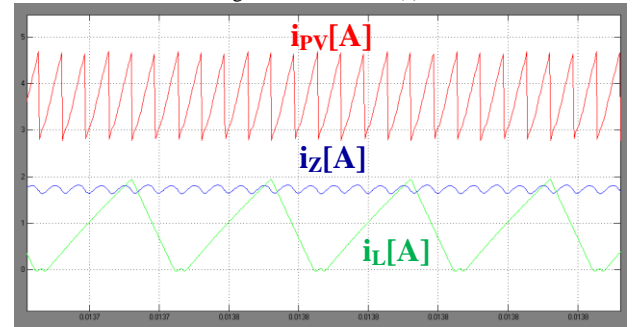


Fig.9. Extended waveforms of inductor current $i_L(t)$, load current $i_z(t)$ and photovoltaic current $i_{pv}(t)$.

VI. CONCLUSION

The simulation results confirm the theoretical assumes. The efficiency of energy conversion of proposed multiphase converter is very high because the output PV energy is continually delivered to the load by means of six phases of multiphase converter. This new concept of proposed converter ensures utilization of the full range of energy supplied from the PV module.

ACKNOWLEDGMENT

The paper has been prepared under support of Slovak grant projects KEGA No. 005TUKE-4/2012, KEGA No.024TUKE-4/2012 and VEGA No. 1/0559/12.

REFERENCES

- [1] Fundamental of photovoltaic materials – 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., Výkonové tranzistory MOSFETa IGBT, Košice: Elfa, 1996. 117 s.
- [4] Kováčová, I. – Kováč. D. - Safeguard circuits of power semiconductor parts. In: Acta Electrotechnica et Informatica. Roč. 3, č. 3 (2003), s. 44-51. - ISSN 1335-8243.
- [5] Kováč, D. - Kováčová, I. - Modelling and measuring of the electronic circuits. Košice: Elfa, 1996, 92 s.
- [6] Kováčová, I. – Kováč. D. – Oetter, J.: Aplikovaná elektronika: návody na cvičenia I. Košice: Akris, 2001, 94 s., ISBN 80-968666-0-5.

Modeling support services by using Modes

¹Jozef DUDIÁK (1st year Ing), ²Pavol HOCKO (2nd year PhD)

¹Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

²Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

¹jozef.dudiak @student.tuke.sk, ²pavol.hocko@tuke.sk

Abstract — this paper describes the process of support services. These services are simulated on interconnected power system with fourteen nodes and four separated areas which are connected together with inter-tie line. On this paper support services are described in details. Each regulation of active power is on separate figure as the output of simulating program Modes. Subsequently, the process of primary regulation is compared on the generator with different performance. The secondary regulation is compared with primary separately. The second part discuss about the power system blackout and the solution how to prevent that.

Keywords—power system, support services, frequency deviation, MODES.

I. INTRODUCTION

Unsustainable development in the executive computing presents various simulating and computing environments, which are able to calculate a steady-state of the power system. The precise definition of different types of faults or changes in network topology allows us to monitor not only the dynamic behavior of the power system but also the individual network elements.[1]

The support services are services that are designed to ensure the provision of system services which is necessary for maintaining the quality of electricity supply and to ensure the operational reliability of the power system.[3]

The frequency is global (system-wide) parameter, which indicates the quality of the electricity. We can maintain it by changing the active power generation or consumption.[2]

Types of control of active power:

- Primary control of active power

Primary control of active power is the response of power turbine to frequency deviations from nominal values. The response must be within a few seconds. Primary regulation participates on the stabilization of frequency deviation and provides it all generators which are operating in the interconnected system and are involved to this regulation.[2]

- Secondary control of active power

Secondary regulation stabilizes the balance between areas to required value and recovery frequency to required value. Secondary control of active power is working in the regulatory area within the time frame of minutes.[3]

- Tertiary control of active power

Tertiary control is used to ensure sufficient power reserve for secondary control. Enlarge this provision is made by

connecting or disconnecting of a new active power source. In the case of overflow the secondary regulation reserve, occurs to disconnection required performance and reverse.[2]

II. IMPLEMENTING COMPUTER PROGRAM MODES FOR MODELING SUPPORT SERVICES

A. Outgoing circuit

For modeling support services has been used predefined project 'LONG' from basic package of program Modes in which we changed the required parameters and a script as needed.

On the figure Fig. 1 we can see the initial scheme of interconnected power system, which consist of four areas that are connected together with inter-tie lines. As the intervention in dynamic simulation of the power system we set a fault on one of the power sources.

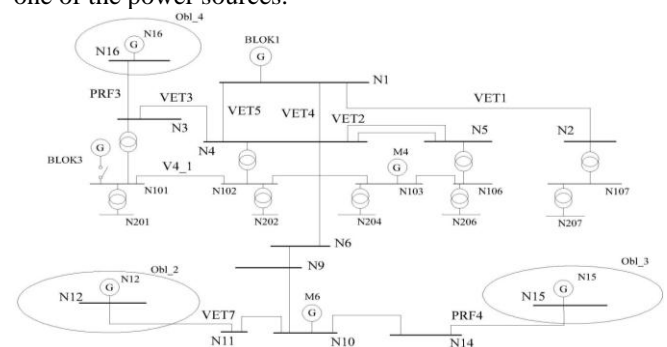


Fig. 1. Initial scheme for modeling support services

B. Simulation the primary control of active power

In the script we set up to turn off unit (BLOK3) by using command 'UNIT' at the time of 10s. The power and speed regulators were left on the default value according to the case. Parameters of nodes, lines and blocks were also kept on the initial values.

After startup the simulation we can see on the output from the graphics the course of the turbine power (real power of the generator) and also courses of the frequency deviation

Mechanical turbine power

Turbine power of Unit M6 (purple) doesn't respond on change of frequency. We used there the coordinated control without frequency corrector and speed control. On the following figure (Fig. 2) is course of the performance on unit

M5, we can see only a temporary response of the turbine power on frequency deviation. Used PI controller adjusts the turbine power, which was temporarily increased by speed controller. Unit M4 has a slower response on change of frequency, causing by a slower opening of the regulatory mechanism of water turbine. We used there the controller to control the power with the frequency corrector that ensures proper function of the primary control.

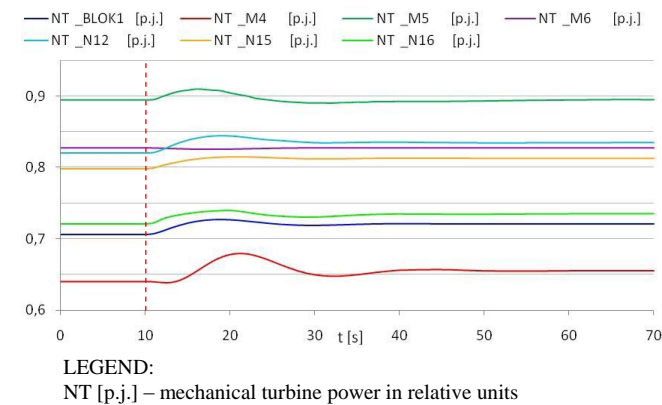


Fig. 2. Time courses of mechanical turbine power units

Unit N16 has the fastest response because the controller operates as the parallel configuration of electronic power regulator with frequency corrector and speed controller. The outputs of both controllers are summed together and ensure quick response on change of frequency. Slower response has the unit N12 because there is used a hydraulic speed regulator in series configuration with electronic power corrector and with the corrector of frequency, which is slower. Even a bit slower is a BLOK1 with regulator which is controlled by an electronic controller of power with a frequency corrector without a speed controller.

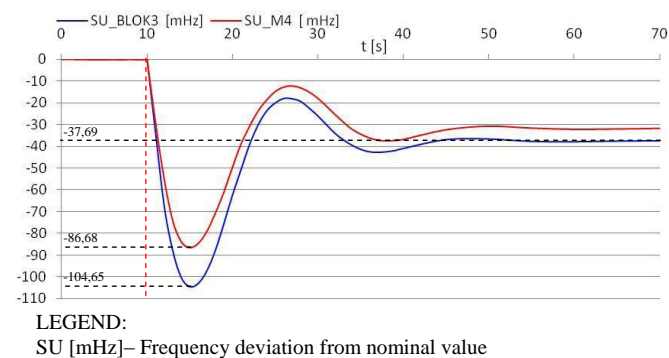


Fig. 3. Process of frequency deviation during the primary control of active power

Frequency deviation

On the figure Fig. 3 we can see the graphical output of frequency deviation from nominal value during outage of two units with different power ranges. Deviation of frequency during failure of unit BLOK3 at the time of 10s in which generator operates with power range of 80-190 MW is shown in blue color. We can see that at the moment of disconnection the unit (reduction of production) frequency decreases strongly.

Dynamic frequency deviation has value $\Delta f_d = -104.645$ mHz. Automatic activation of the primary control immediately after the outage adjust the quasi-stationary

frequency deviation to this value $\Delta f = -37.688$ mHz. Outage of unit with lower power (M4) is shown on figure Fig. 3 with red color. This outage occurs at the time of 10s too. Unit M4 is composed of two generators with the power range of turbines from 20 to 130 MW. Dynamic frequency deviation has value $\Delta f_d = -86.678$ mHz and quasi-stationary frequency deviation has value -31.7 mHz. From figure we can see that if it is outage of unit with greater power the frequency deviation is then greater.

C. Simulation the secondary control of active power

Before we start to simulate the secondary regulation of active power we set the activation of automatic with the command 'AUTO' and then assigning an object 'LFC' at time 0s. This automatic performs the regulation of power control and frequency. To set the failure we must use command 'UNIT' to turn off one of the units (BLOK3) at the time of 10 seconds.

Power and speed regulators, parameters of nodes, branches and units had remained unchanged. When we start simulation the output of graphic shows the courses of turbine performance, frequency deviation and power flows (balance) between control areas.

Units operating in the areas where the failure of power unit (BLOK3) was, provides the secondary control and assumes the load of failed out unit. These units adjust the balance and the frequency deviation to set point. Other units that operate in the areas 2-4 provide only emergency assistance to areas where was the failure, according to principle of solidarity and do not provide a secondary power control. Secondary power control provides only units that operate in areas where the outage occurs. The power of turbines on these units is changed by primary regulation when it comes to removing frequency deviation it returns the power of turbines on these units to default value. The balance between production and consumption ensures each system alone.

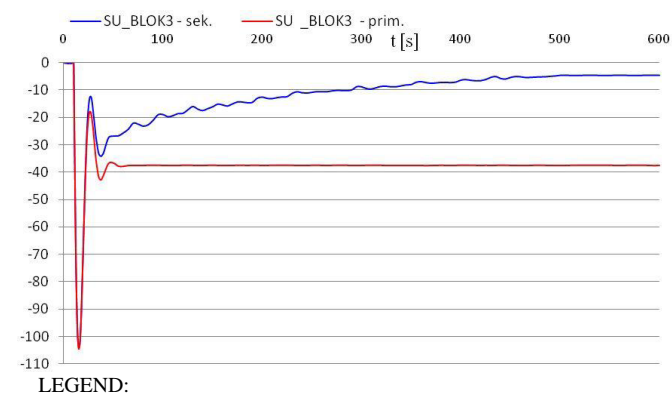


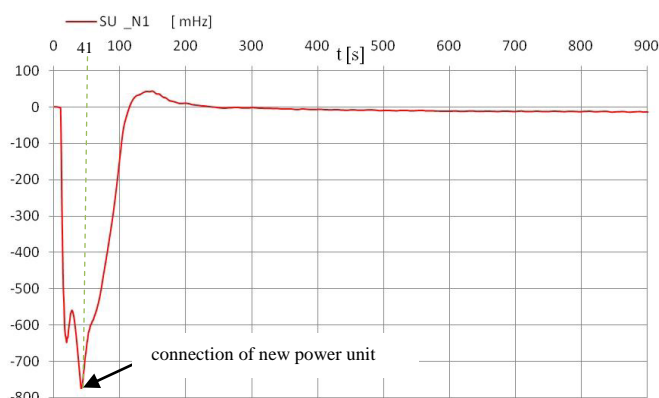
Fig. 4. Process of frequency deviation during the secondary control of active power

On the figure Fig. 4 we can see comparison of frequency deviation with activated (blue) and deactivated (red) secondary power control. At the time of 40s occurs to activation of the secondary power control. Due to central regulators which evaluate the control deviation ACE is regulating the frequency deviation. From output of simulation we can see that at the time of 600s was the frequency deviation balanced to the value -4 mHz, which can be considered as good result of secondary regulation, because

this regulation should complete its activation within 15 min. At the same time was balanced the power balance of individual areas, which confirms the correct functioning of the secondary regulation. On the figure Fig. 4 we can see the difference in the operation of primary and secondary power regulation. During the primary regulation, the deviation of frequency was stabilized at quasi-stationary deviation. The effect of secondary regulation regulates the frequency deviation to zero value.

D. Simulation the tertiary control of active power

To simulation of tertiary regulation, we had assigned a new unit (BLOK2) to our model of power system. This unit consists of generator with power range from 450 to 750 MW. The initial state of this unit we set to OFF (set 0). In the scenario we have activated the secondary power control. With command 'LOAD', which is used to sudden changes in load, we have increased the load from 0 to 400 MW at nodes N201 and N202. With another command 'SYNC' at the time of 40s we have launched a synchronization of new unit to the interconnected system. At the time of 41s we have joined this new unit (BLOK2). We must set the outputs from graphics to see frequency deviation in the interconnected system and to see the power of unit N1.

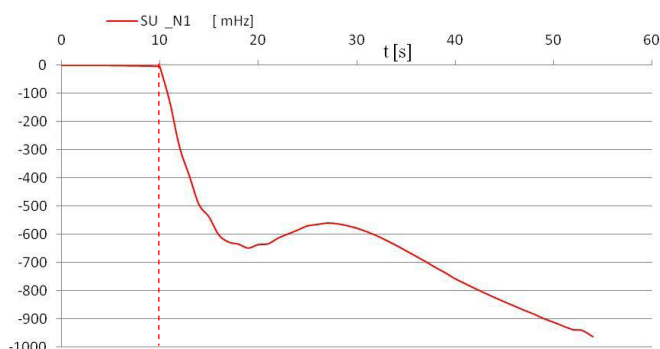


LEGEND:

SU [mHz]– Frequency deviation from nominal value

Fig. 5. Frequency deviation during connecting new power unit

On the figure Fig. 5, we can see the progress of the frequency deviation. Sharp increase of the frequency deviation at the time of 10s occurs load increasing. The frequency deviation is so large that this rising cannot stop the primary power control. The turbine power of all units is rapidly increasing. At the time of 41s occurs the connection of new unit. If the operators of the system did not respond correctly and did not give the order to connect a new unit and thus increase of secondary regulatory reserve, there would be a breakdown of the interconnected system, which can be traced on Fig. 6. After connecting of a new power source we can see balancing of frequency deviation. This new unit covers the performance deficit caused by a sharp increase in load.



LEGEND:

SU [mHz]– Frequency deviation from nominal value

Fig. 6. Frequency deviation during braking up the power system

III. CONCLUSION

Support services are the basis of managing power systems. The primary control of active power responds automatically to change the frequency from the set value, thus ensuring stability of the system at any time. If the primary control did not respond to these changes, will occur failures in the interconnected system of various kinds, whether sudden breakdown of the system or failures in synchronism. Every such failure causes loss of production and also restrictions on supplies to customers.

The primary control of active power is closely associated with the secondary control of active power, as the primary control only stabilizes the frequency on the quasi-stationary value, secondary control by changing the active power, adjust it to set value in required tolerance. This secondary control operates automatically. Dispatcher can use various resources to ensure this stability whether it is connecting or disconnecting of new sources or disconnecting the loads. Many of these resources shall be included in tertiary regulation of active power, which must respond within the required time according to requests sent from the control room. Each machine providing the tertiary power control must be regularly tested and must be ready to roll on full alert at all times under the conditions set by transmission system operator (ENTSO-E). The aim of support services simulation was to clarify the operation of primary, secondary power control and also highlight the importance of tertiary power control. Courses of all power controls are in accordance to theoretical assumptions, all time and operating parameters was observed.

REFERENCES

- [1] DUDIÁK, Jozef.: Bakalárska práca, Podporné služby v elektrizačnej sústave Slovenskej republiky.
- [2] KOLCUN, Michal et al.: Prevádzka elektrizačnej sústavy. Košice: TU, 2007. 306 s. ISBN 978-80-8073-837-2.
- [3] SEPS a.s.: Prevádzkový poriadok prevádzkovateľa prenosovej sústavy [online]. Dostupné na internete: <http://sepsas.sk/seps/Dokumenty/2010_PP_SEPS_as_BA_1.pdf>.
- [4] MASLO, Karol.: Popis modelování přechodných dějů v ES programem MODES v.2.2./8.

Printing Performance of Solder Paste Dependent on Temperature and Density

¹Michal KRAVČÍK (3rd year),
Supervisor: ²Alena PIETRIKOVÁ

^{1,2}Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic

¹michal.kravcik@tuke.sk, ²alena.pietrikova@tuke.sk

Abstract—The study of this paper is oriented to identification of influence of various density as well as temperature influence on printing performance of solder pastes with the aim to identify solder paste that produce a required stencil print. This study concerns on rheological characterization of such three pastes. Paste P1 is commercial solder paste with 88,5% content of metal by the weight with standard rheological behavior and with standard stencil printing performance. Paste P2 was formulated with a higher metal content (90% of metal weight). And Paste P3 (91,5% of metal weight) was produced for the comparison with previous two types. Two different types of test were conducted in the study: viscosity test was used to identify any shear-thinning behavior of this non-Newtonian fluid and printing quality tests with solder paste inspection. The aim of this paper was analyze of sensitivity of solder paste to the change of temperature in the range from 20 to 30 °C as well as density from 88,5% to 91,5% of metal content on printing quality.

Keywords— printing performance, solder paste, rheological behavior, solder paste inspection,

I. INTRODUCTION

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. The science of rheology is concerned with how materials respond to applied forces and deformations. The understanding of rheological behavior is of paramount importance in many practical applications such as inks, oils, paints, food and cosmetics. Rheology of solder pastes is a vital factor as it directly affects the printing characteristics and post-print performance of the solder paste. Solder pastes (Pb-free) are concentrated suspensions of solder particles (Sn-Ag-Cu) in a flux system. The flux system in solder paste is a combination of solvents, fluxing chemicals, thickeners, and tackifiers. The knowledge of the rheological properties of solder pastes is also required for the design, selection and operation of the equipment involved for mixing, storage and printing of solder pastes. 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.

Now days trends of continued miniaturization of electronic

components is one of the way to accomplish perfect solder joint, is decreasing of flux content of solder paste, mainly if Vapour Phase Soldering is expected. Reducing of flux volume should reduce solder volume difference before and after a solder melting. This volume difference can cause of non welcome gap between solder and solder pad on component.

The goal of the work is to examine how increasing metal content influences rheology properties of SnAgCu solder pastes and accomplish good printing quality with higher metal content pastes. The aim of this paper is analyze the sensitivity of solder paste on various temperature in the range from 20 to 30 °C on printing quality as well.

II. EXPERIMENTS

Our experiments were oriented to identification of influence of various densities as well as temperature influence on printing performance of solder pastes with the aim to identify the technological condition in which solder paste has very good stencil print quality. Three various type of solder pastes were used for rheology analyze as well as for SPI identification.

A. Material and Sample Preparations

One commercial grade solder paste (Paste P1) was modified by increasing of metal content to 90% of metal weight (Paste P2) and 91,5% of metal weight (Paste P3). All composition were based on 96.5Sn-3Ag-0.5Cu, flux type M14, melting temperature 217 – 220°C, diameter of particles were 25 – 45 µm. SnAgCu powder particles characterised by the identical physical properties in compare with commercial paste was admixed.

B. Viscosity Tests

Solder paste is pseudoplastic Non-Newtonian fluid, that's mean apparent viscosity decreases with increased stress. When solder paste is moved by the squeegee on the stencil, the physical stress applied to the paste causes the viscosity to break down, thinning the paste and helping it flow easily through the apertures on the stencil. When the stress on the paste is removed, it regains its shape, preventing it from flowing on the circuit board. For viscosity tests was used very accurate rheometer system Anton Paar 501 with a cone - plate (CP) measuring geometry of 25 mm diameter and 0,5 mm gap was used, as shown in the Fig. 1.

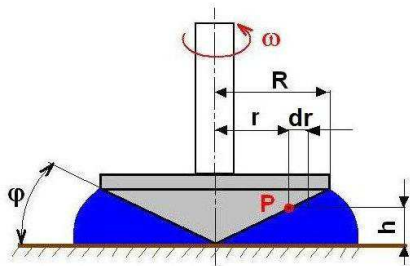


Fig. 1: Cone - plate measuring geometry (R - cone radius, ω - angular velocity, P - arbitrary point on cone, ϕ - bevel angle, h - y coordinate of P point, r - x coordinate of P point)

All rheological measurements were conducted at machine temperatures from 20°C to 30°C (accuracy $\pm 0.1^\circ\text{C}$) with the temperature being controlled by a Peltier-Plate system. This temperature simulate realistic condition in manufacture hall in production assemble line. And area of shear rate $\dot{\gamma}$ was from 1s^{-1} to 25s^{-1} , which covers all shear rates inside of solder paste, during stencil printing process.

C. Printing Quality Tests

Tests for quality solder pastes P1, P2 and P3 were made because of the need to ensure print quality with high repeatability. Each paste was applied the same procedure on 5 PCB test sample segment, while the temperature of printing simulate real conditions in a factory building, that is ranged at 24 °C, 28 °C and 30 °C. At these temperatures were also carried out experiments to analyze the rheological properties. This test measured the height of paste deposited and total volume applied to each solder pads and also measurement of deposition defects. Printing quality tests were performed at two different solder pads segments. First segment has pitch size 300 μm and second segment only 150 μm .

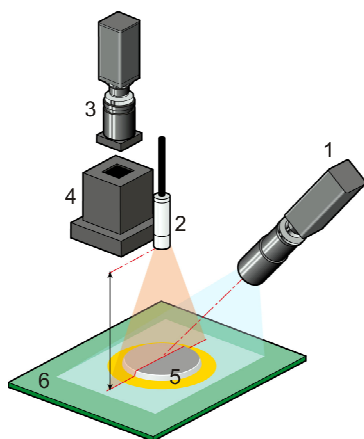


Fig. 2: Principle of measuring values (position, shape, height and etc.) of deposit solders paste volume at solder paste inspection (SPI) system. 1 - Laser source, 2 - laser detector, 3 - operator camera, 4 - optics, 5 - deposited solder paste, 6 - printed circle board.

For measurement of deposited solder paste volume and height was used 3D solder paste inspection system (SPI). The principle of 3D solder paste inspection process is shown in the Fig. 2. Innovative 3D laser triangulation sensor profiles real shape of features and also provides highly robust 3D data against considerable variations of PCB color and finishing condition, solder material, paste shape, and warpage. Intelligent vision algorithm applied to the high quality 3D data detects every kind of solder paste defects. The measurable

items are height, area, volume, and positional offset that deviate from tolerance limits. And also the algorithm detects bridged pads in a very reliable manner. The 3D shape modeled on the 3D viewer gives user other meaningful information about the printing process. The shapes of solder pastes are tightly coupled to the screen printer parameters, such as the squeeze speed, squeeze pressure, snap off, etc. The 3D images on the viewer can be enlarged, translated, rotated to help operator's further detailed analysis. And cross sectional profiles of a pad are displayed on the graph. Example of solder paste deposition defect is in the Fig. 3.

Stencil printer performance is typically quantified by measuring the transfer efficiency percentage and standard deviation of the paste deposition, where 100% would mean that the printer paste profile matches the calculate volume of the aperture. It is interesting to note that transfer efficiency percentage varies on a typical SMT board between 20 to 130% (lit).

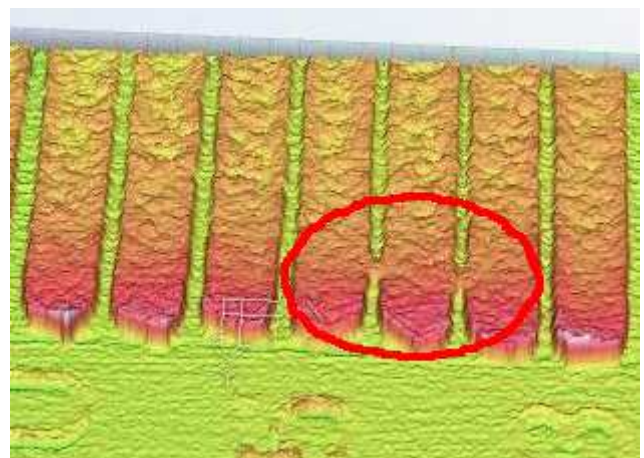


Fig. 3: Example of deposition defect, called bridging. Picture is taken by 3D SPI system.

In the fig. 4 is ideal deposition of sample Paste P2 in compare with defect deposition of sample P1 realized at the same temperature 28°C. Walls of defected deposited paste have different height of solder paste on opposite sides, typically in direction of squeegee motion. It could be named as one-side corner warpage. Solder paste P2 with higher density shows on warpage compensation that is typical for higher printing volume accuracy.

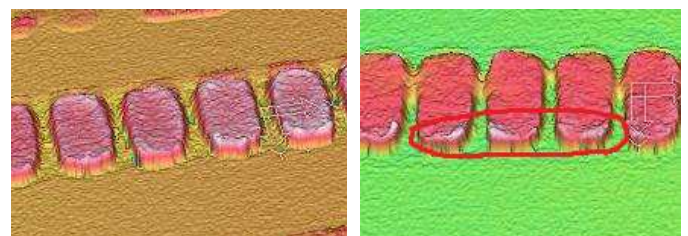


Fig. 4 Comparison of good printed solder paste (left) and solder paste with defect deposition call one-side corner warpage.

III. RESULTS

A. Viscosity Tests

Results of viscosity flow curves are in the Fig. 5, and that curves shown pseudoplastic behavior of solder paste as well. We can see high difference between solder pastes viscosity at low shear rates. At low shear rates from 1 to 5s^{-1} are different

of viscosity values more than 50% between samples P1 and P3. At higher shear rates (more than 5 s^{-1}) are differences of viscosity values less than 30% between P1 and P3 and less than 18% between P1 and P2.

In this context we can see with higher content of metal weight increase the approved viscosity value in the same temperature. Very important result influenced from this experiments: amount of metal in the solder paste doesn't influence pseudo-plastic behavior.

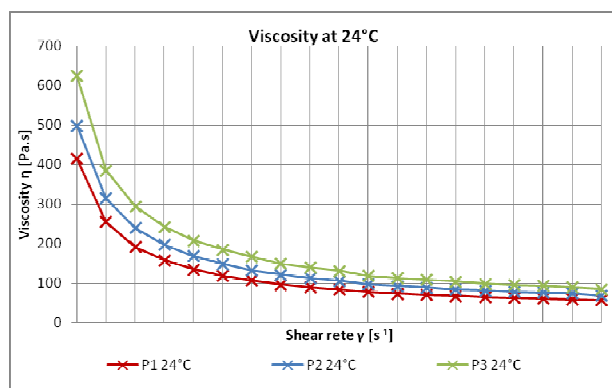


Fig. 5: Flow curves of pastes P1, P2 and P3 at temperature 24°C.

In the Fig. 6 are compilations between viscosities of all pastes in the dependent on temperature. We can see that the same value of viscosity can occur for two solder paste samples at different temperature. For example viscosities P1 at 22°C and P2 at 28°C are identical. There is a postulate that if viscosities of two different paste samples have identical viscosity, they should have an identical printing quality results. All of solder paste samples have almost linear decreasing of viscosity values with temperature rising. In this figure we can see that very small change of temperature (from 20 – to 30 °C) bring important decreasing in viscosity (from 176 to 135 Pa.s for sample P3). An angle of decreasing gradient of viscosity is identical for all three samples. From this Fig. results fact that temperature sensitivity of solder paste viscosity doesn't depend on metal content. Temperature sensitivity depends on flux behavior, which is identical on all used samples.

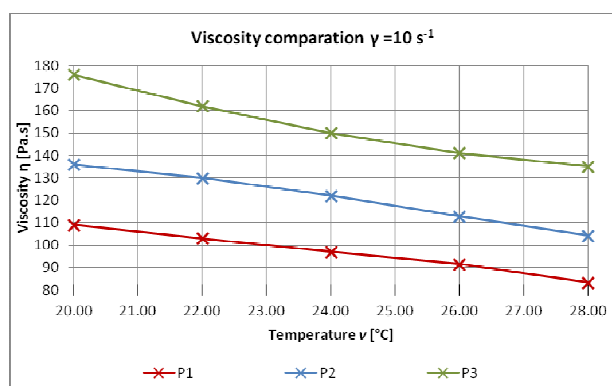


Fig. 6: Linear decrease of viscosity values with temperature rising for all three solders paste samples.

In the Fig. 7 is 3D graph of viscosity established from measured viscosity values of all solder paste samples. Via this data is build 3D surface of viscosity values with dependency on temperature and metal content. Validity of this fig. is

correct only if flux type and metal particle size are the same like in these experiments. 3D graph of viscosity with dependency on temperature and metal content of solder paste is ideal for visualization of solder paste behavior in the dependence of density and temperature.

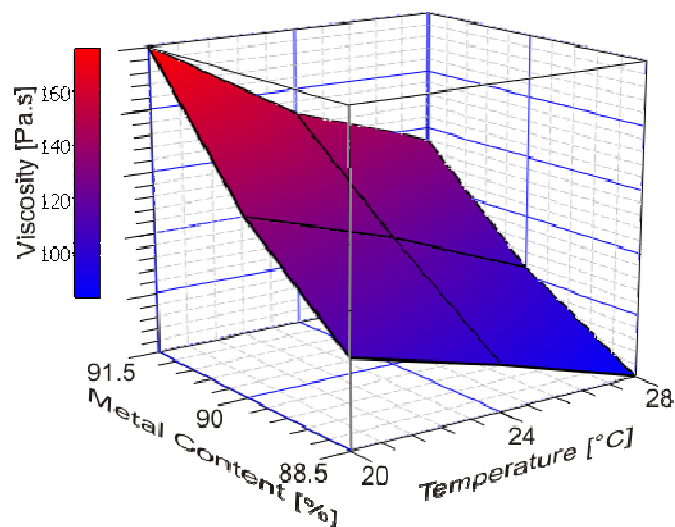


Fig. 7: 3D graph of viscosity with dependency on temperature and metal content of solder paste.

B. Printing Quality Tests

Printing quality tests on segment with pitch 300 μm and 150 μm have few results. In the Fig. 8 and also the Fig. 9 are percentages of errors (solder paste deposition defects) from repeated printing tests.

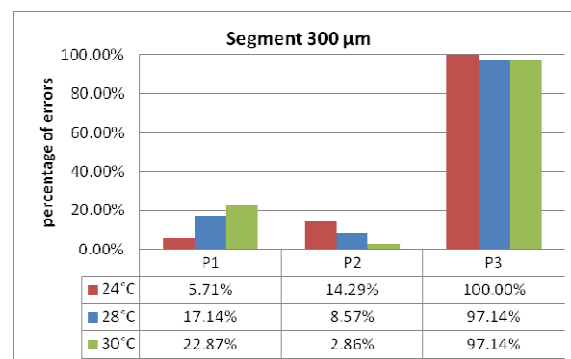


Fig. 8: Percentage of printing errors for the segment 300μm.

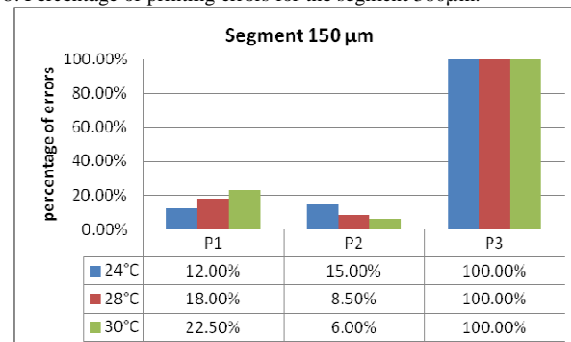


Fig. 9: Percentage of printing errors for the segment 150μm.

High number of defects (more than 95% of errors) of sample P3 means that this sample with the highest metal content (91,5%) is not able to print with standard stencil printing technique. Density of such paste is too high and for SMT

application is not proper. We can say that 91,5% is critical content of metal in solder paste. Results for sample P1 means that raising of temperature worsens printing ability of commercial grade solder paste. Results of sample P2 show us an improvement of printing quality with rising temperature. Like in viscosity tests sample P1 at 24°C and sample P2 at 30°C have very similar are percentages of errors on both tested segments. This result is the biggest contribution influenced from this printing quality analyze. Paste P3 is not proper for printing because of very high value of its density (viscosity) in spite of the higher temperature. From this point of view it is important to say that ideal behavior of solder paste for printing quality has paste P2 in the whole temperature range.

IV. CONCLUSION

In this paper we have analyzed the sensitivity of solder paste to the change of temperature in the range from 20 to 30 °C as well as to the density from 88,5% to 91,5% of metal content on printing quality. Higher content of metal weight increase the approved viscosity value in the same temperature. Amount of metal in the solder paste doesn't influence pseudo-plastic behavior. If viscosities of two different paste samples have identical value, they should have an identical printing quality results what was confirmed in printing quality tests. All of solder paste samples have almost linear decreasing of viscosity values with temperature rising. Sensitivity to temperature is not influenced by the density of solder paste. Sensitivity on temperature depends on flux behavior, which is identical on all used samples. Ideal behavior of solder paste printing process has paste P2 in the whole temperature range. The best printing quality was recorded for sample P2 at temperature 30°C. This paste has bigger density and is comparable (or better) with commercial paste P1.

ACKNOWLEDGMENT



This paper was developed with support of the project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií" v oblasti automobilovej elektroniky", ITMS 26220120055, that is co-financed from Structural Funds EU ERDF within Operational programme Research and Development OPVaV-2009/2.1/03-SORO and preferred axis 2 Support of Research and Development.

REFERENCES

- [1] Cheng, D.C.H.; Phenomenological Characterisation of the Rheological Behavior of Inelastic Reversible Thixotropic and Antithixotropic Fluids. 1965.
- [2] Malik, S., Ekere, N.N., Seman, A.; Modeling the Structural Breakdown of Solder Paste Using the Structural kinetic Model. 2009. ASM International DOI: 10.1007/s11665-009-9448-0 1059-9495/
- [3] Dušek, M., Ling, Z., Hunt, Ch.; Rheology Testing of Solder Pastes and Conductive Adhesives. NPL MATC(A) Report 109, May 2002.
- [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, 2009.
- [6] Y. Tian, Yan C. Chan, Senior Member, IEEE, J. K. L. Lai, and Sally T. F. Pak.: The Effect of Solder Paste Viscosity on Porosity and Mechanical Properties of Surface Mount Solder Joints. IEEE TRANSACTIONS ON COMPONENTS, PACKAGING, AND MANUFACTURING TECHNOLOGY—PART B, VOL. 20, NO. 2, MAY 1997
- [7] R. Durairaj, S. Mallik, A. Marks, *M. Winter, *R. Bauer and N.N. Ekere.: Rheological Characterisation of New Lead-Free Solder Paste Formulations for Flip-Chip Assembly; 2006 Electronics Systemintegration Technology Conference Dresden, Germany; IEEE 1-4244-0553-1/06/
- [8] JEONG HWAN KIM, MUNETAKE SATOH and TOMOHIRO IWASAKI: Rheological properties of particle-flux suspension paste

A comparison the effectiveness of induction motor's diagnostic methods based on spectra analysis of current and instantaneous power signals

¹Dmytro MAMCHUR, ²Andrii KALINOV, ³Ján MOLNÁR

^{1,2} Institute of Electromechanics, Energy Saving and Control Systems,
Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine

^{1,2}scenter@kdu.edu.ua ³jan.molnar@tuke.sk

Abstract – In the article the comparison of effectiveness the induction motors fault detection methods, based on experimental results analysis, was done. Researches showed advantages of the method, based on total 3-phase instantaneous power spectra analysis.

Keywords – Diagnostics, induction motor, instantaneous power, spectra analysis.

I. INTRODUCTION

Nowadays induction motors (IM) are the most widely used consumers of electric energy. According to last researches, they consume over 80 % of the total amount of electric power [1]. Induction motors widely used because of the simplicity of their construction, reliability and relatively low cost. Creation of speed-controlled electric drive systems resulted in increased usage of IM because of the simplicity their implication into such systems. However, different damages could be caused to IM parts under operating conditions. They could lead to untimely motor failing. Every year as a result of failures about 20-25 % of induction motors fails. Thus, there is an important task of creation reliable IM diagnostic system. In many manufactures sudden failure of the motor could lead to shutdown the technological processes. Operation of electric motors under unsatisfactory technical condition results both in direct financial losses, and to considerable (about 5-7 %) indirect expenses of the electric energy. On this and other reasons there is a necessity of IM diagnostics under operating conditions. The most convenient and reliable IM diagnostic methods are based on the electric signal analysis. There are two convenient methods for on-line IM diagnostic – current spectra analysis and instantaneous power spectra analysis. First one needs only one phase current signal for analysis. So this method is attractive of its measuring simplicity. Second one needs data of three phase currents and voltages. This method is more complicated, but gives more reliable results. So the aim of presented work is comparison of IM diagnostic methods based on current spectra and instantaneous power spectra analysis.

II. MOTOR CURRENT SIGNATURE ANALYSIS (MCSA)

MSCA is a very popular diagnostic method because of simplicity of signal recording under operation mode. There is a number of works devoted to using MSCA as a medium for detection of stator windings short-circuits, rotor unbalances and rotor bar breaks, and also bearings defects [2]. The idea of this method is corresponding IM defects to certain harmonics in current spectra as it is shown follow.

A. Broken rotor bars

In case of broken bars, the disturbance induces fundamental sidebands of the supply frequency and of the winding factor into the current signal [2]:

$$f_{bb} = f_s \left[\frac{k}{p} \pm s \right] \quad (1)$$

where f_s is the supply fundamental frequency,

k is an integer number,

p is the number of pole pairs,

s is the motor slip,

$k/p = 1, 5, 7, 11, \dots$

B. Air gap eccentricity

There are three types of air gap eccentricity: static, dynamic and mixed. All these eccentricities lead to appearance two types of unique current frequencies: around fundamental harmonic (f_{eccen_fun}) and around principle slot harmonic (f_{eccen_prin}) [2]:

$$f_{eccen_fun} = f_s \pm k f_r, \quad (2)$$

where $f_r = \frac{1-s}{p}$, where s is the motor slip;

$$f_{eccen_prin} = \left[R \pm n_d \left(\frac{1-s}{p} \right) + \eta \right] f_s \quad (3)$$

where R is the number of rotor slots,

n_d is an integer due to dynamic eccentricity,

η is the time harmonics present in the motor supply.

C. Bearings damage

Bearings damage leads to vibrations. Since bearings support rotor, their vibrations, in turn, leads to air gap fluctuations. These variations lead to stator current

frequencies given by [2]:

$$f_{brg} = f_s \pm m f_{i,o}, \quad (4)$$

where $m = 1, 2, 3, \dots$, and $f_{i,o}$ is one of the characteristic vibration frequencies which are based upon the bearing dimensions:

$$f_{i,o} = \frac{n}{2} f_v \left[1 \pm \frac{bd}{pd} \cos \beta \right],$$

where n is the number of bearing balls,

f_v is the mechanical rotor velocity in hertz,

bd is the ball diameter,

pd is the bearing pitch diameter,

β is the contact angle of the balls on the races.

As it could be seeing from equations (1)–(4), all reviewed damage types could lead to similar or identical harmonic components. Thus, it is additional features demanded to determine certain damage type. Also it has to be mentioned that electrical distortions and influence of supplying voltage low quality could lead to appearance of harmonics in current signal on the same frequency as a fault harmonics. This may lead to wrong diagnosis. To eliminate such shortcomings of this method the additional analysis by vibrations [3] or more complicated mathematical apparatus for analysis [4, 5] are used. However, even this additional analysis does not protect from diagnostic mistakes when low-power IMs with significant influence of voltage unsmoothness are under analysis. This fact is especially significant for IM fed by low-voltage supply of industrial plants.

III. INSTANTANEOUS POWER SPECTRA ANALYSIS

The instantaneous power spectra analysis allows avoiding shortcomings of above mentioned methods [6-8]. Instantaneous power spectra analysis allows both detection of fault presence and estimation of damage level by analysis of proper harmonic value. Thus, it allows one to make estimation of the energy of fault and the correlation of this energy to additional damage of IM parts under influence of additional vibrations caused by proper harmonic. Moreover, the instantaneous power spectra analysis allows analyzing of IM operation modes under significant nonlinearity, when it is incorrect to use superposition principle for current harmonics. Also, instantaneous power analysis is more reliable, it is less dependent on noise, and gives additional harmonic components for analysis [6-8].

The instantaneous power is defined as:

$$p(t) = u(t)i(t)$$

where $u(t)$ is the phase voltage;

$i(t)$ is the input phase current.

In case of a healthy motor running with a constant speed and fed from ideal supply, the expressions of the phase voltage $u(t)$, phase current $i(t)$ and instantaneous power, are following [1]:

$$u(t) = \sqrt{2} U_1 \cos(\omega t);$$

$$i(t) = \sqrt{2} I_1 \cos(\omega t - \varphi);$$

$$\begin{aligned} p(t) &= u(t)i(t) = 2U_1 I_1 \cos(\omega t) \cos(\omega t - \varphi) = \\ &= 2U_1 I_1 \cos(\omega t) \left[\cos(\omega t) \cos(\varphi) + \right. \\ &\quad \left. + \sin(\omega t) \sin(\varphi) \right] = \\ &= U_1 I_1 \cos(\varphi) + U_1 I_1 \cos(\varphi) \cos(2\omega t) + \\ &\quad + U_1 I_1 \sin(\varphi) \sin(2\omega t), \end{aligned} \quad (5)$$

where U_1, I_1 are RMS values of phase voltage and current, respectively;

$\omega = 2\pi f$ is the angular frequency,

where f is the supply frequency;

φ is the motor load angle.

In difference to current spectra, which contain only the fundamental component at the frequency f , the instantaneous power spectra has an average power component $U_1 I_1 \cos(\varphi)$ and fundamental component at frequency $2f$.

In order to make comprehensive analysis of IM defects, it is necessary to analyze total instantaneous power of three phases, which is the sum of phase instantaneous powers:

$$p_{tot}(t) = u_A(t)i_A(t) + u_B(t)i_B(t) + u_C(t)i_C(t).$$

Total three phase instantaneous power contains more diagnostic information. It allows us to analyze not only defects which causes phase signals modulations, but also the defects which caused by motor or supply asymmetry and which leads to phase signals asymmetry. In case of symmetrical motor, signal of total three phase instantaneous power contain only DC component. Thus, every kind of motor fault or drive system asymmetry leads to appearance unique harmonic components which could be used for certain detection of fault type.

A. Rotor bar breaks

As it was mentioned above, rotor bar break causes sinusoidal modulations of the stator current. By analogy to [8], modulated phase current can be expressed as:

$$\begin{aligned} i_m(t) &= i(t) [1 + I_m \cos(2\pi f_{bb} t)] = \\ &= i(t) + \frac{\sqrt{2}}{2} I_1 I_m \left[\cos(2\pi(f - f_{bb})t - \varphi) + \right. \\ &\quad \left. + \cos(2\pi(f + f_{bb})t - \varphi) \right], \end{aligned} \quad (6)$$

where I_m is the modulation index,

f_{bb} is the modulating frequency,

s is the motor slip.

According to expression (6), phase current spectra, in addition to fundamental component, contain two sideband components at frequencies $f - f_{bb}$ and $f + f_{bb}$.

Expression for modulated phase instantaneous power is the following:

$$\begin{aligned} p_m(t) &= i_m(t)u(t) = \\ &= p_0(t) + \frac{1}{2} I_1 I_m U_1 \cos[\pi(2f - f_{bb})t - \varphi] \\ &\quad + \frac{1}{2} I_1 I_m U_1 \cos[\pi(2f + f_{bb})t - \varphi] \\ &\quad + I_1 I_m U_1 \cos(\varphi) \cos(f_{bb} t) \end{aligned} \quad (7)$$

This expression shows that phase instantaneous power spectra, besides DC component $p_0(t)$ and two sideband components at frequencies $2f - f_{bb}$ and $2f + f_{bb}$, contains an additional component $I_1 I_m U_1 \cos(\varphi) \cos(f_{bb} t)$ at the

modulation frequency f_{bb} , which is an additional diagnostic parameter.

In case of symmetrical drive system there is compensation of harmonic components in 3-phase instantaneous power spectra. Thus, it contains only DC component and component at frequency $2f + f_{bb}$.

Any kind of drive system asymmetry leads to appearance additional harmonic components in 3-phase instantaneous power spectra at the modulation frequency f_{bb} , sideband component at frequency $2f + f_{bb}$, and fundamental component at frequency $2f$.

B. Air gap eccentricity

In case of air gap eccentricity, current frequencies expressed by (2) leads to modulation of phase current:

$$i_{eccen}(t) = i(t) + \frac{\sqrt{2}}{2} I_l \sum_{k=1}^K \left[I_{e1k} \cos(2\pi(f - kf_r)t - \alpha_{e1k}) + I_{e2k} \cos(2\pi(f + kf_r)t - \alpha_{e2k}) \right], \quad (8)$$

where I_{e1k}, α_{e1k} are the current amplitudes and the initial phase angle for frequencies $f - kf_r$,

I_{e2k}, α_{e2k} are the current amplitudes and the initial phase angle for frequencies $f + kf_r$.

In this case the expression for phase instantaneous power of motor operating under air gap eccentricity is the following:

$$p_{eccen}(t) = i_{eccen}(t)u(t) = p_0(t) + \frac{1}{2} I_l U_l \sum_{k=1}^K \left[I_{e1k} \cos[\pi(2f - kf_r)t - \alpha_{e1k}] + I_{e2k} \cos[\pi(2f + kf_r)t - \alpha_{e2k}] + I_{e1k} \cos(\alpha_{e1k}) \cos(kf_r t) + I_{e2k} \cos(\alpha_{e2k}) \cos(kf_r t) \right]. \quad (9)$$

Air gap eccentricity leads to appearance in phase power spectra sideband components at frequencies $2f - kf_r$ and $2f + kf_r$, and additional harmonic components at frequencies kf_r .

C. Bearings damage

The vibrations caused by bearings damage leads to current modulation at frequencies according to (4):

$$i_{brg}(t) = i(t) + \frac{\sqrt{2}}{2} I_l \sum_{k=1}^K \left[I_{b1k} \cos(2\pi(f - kf_{brg})t - \alpha_{b1k}) + I_{b2k} \cos(2\pi(f + kf_{brg})t - \alpha_{b2k}) \right],$$

where I_{b1k}, α_{b1k} are the current amplitudes and the

initial phase angle for frequencies $f - kf_{brg}$,

I_{b2k}, α_{b2k} are the current amplitudes and the initial phase angle for frequencies $f + kf_{brg}$.

By analogy to (9) the instantaneous power of motor operating with the bearings damage is the following:

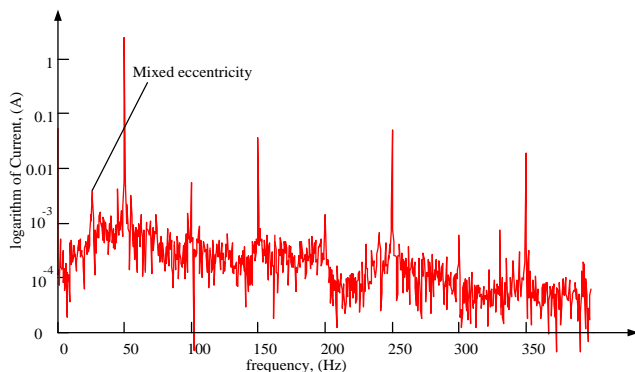
$$p_{brg}(t) = i_{brg}(t)u(t) = p_0(t) + \frac{1}{2} I_l U_l \sum_{k=1}^K \left[I_{b1k} \cos[\pi(2f - kf_{brg})t - \alpha_{b1k}] + I_{b2k} \cos[\pi(2f + kf_{brg})t - \alpha_{b2k}] + I_{b1k} \cos(\alpha_{b1k}) \cos(kf_{brg} t) + I_{b2k} \cos(\alpha_{b2k}) \cos(kf_{brg} t) \right].$$

Bearings damage leads to appearance in phase power spectra sideband components at frequencies $2f - kf_{brg}$ and $2f + kf_{brg}$, and additional harmonic components at frequencies kf_{brg} .

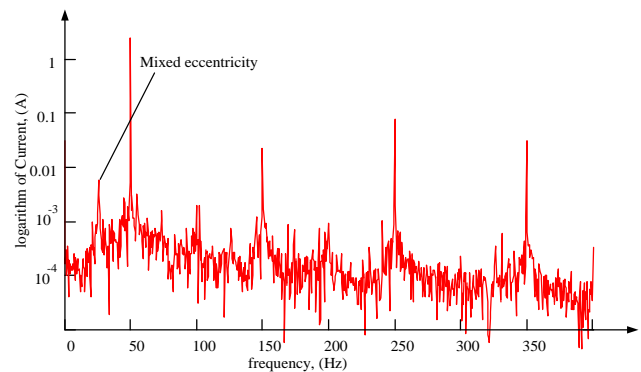
IV. EXPERIMENTAL RESULTS' ANALYSIS

To reach the aim of this work, a series of experiments was done. Three identical induction motors of type AHP80B4Y2, 1.5 kW, were used for testing. These motors were artificially damaged with three most frequently caused damage types: stator winding unsymmetry, rotor bar breaks and rotor eccentricity. Currents and voltages of phases were measured both under idle mode and full load mode, and then they were analyzed. Analysis results leads to the following conclusions (fig.1, 2).

All tested motors have basic mixed eccentricity, caused of motor disassembling and assembling operations. Further analysis showed, that both methods could be used for detecting the different motor damages types, but amplitude values of current spectra harmonics related, for example, to stator unsymmetry, are too small (fig.1, b, d). Thus, in order of incipient fault, they could be wrong detected as a noise harmonics. Moreover, this method is unavailable to detect motor or supply asymmetry. In difference to this method, total 3-phase spectra analysis allows operate with clearly visible harmonics (fig. 2, b, d). Moreover, this method allows get big number of additional harmonic components related to each damage type, and also allows detect asymmetry of electric drive system. This feature also allows avoid wrong diagnosis.

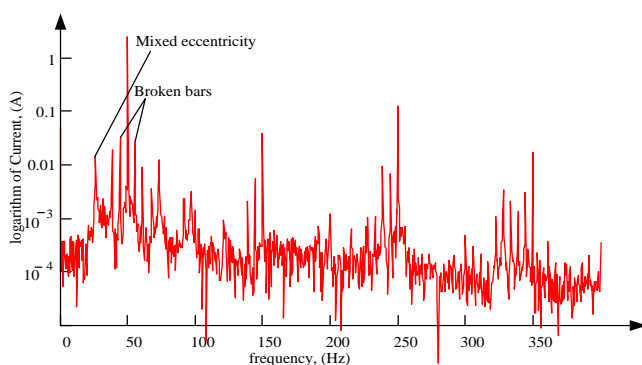


a) Healthy motor



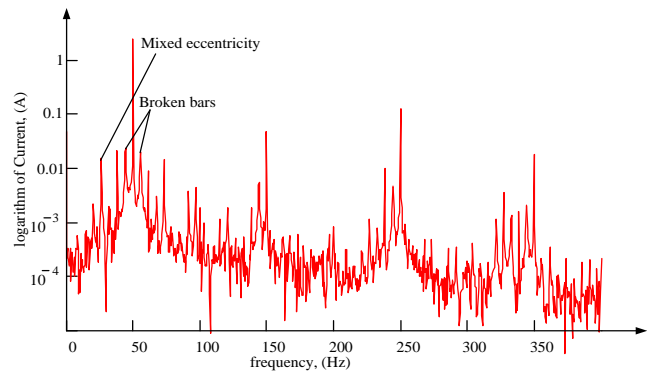
b) 14% stator phase unsymmetry

Fig. 1 Phase current (beginning)

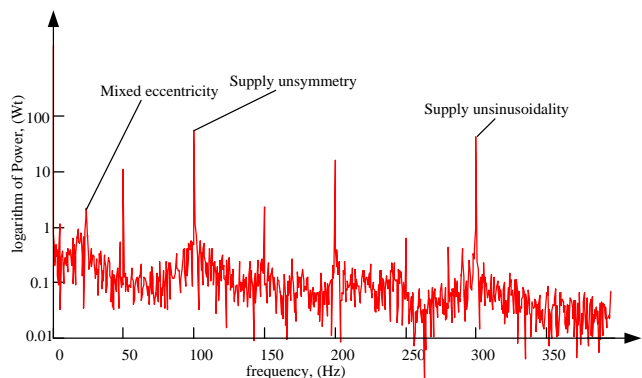


c) 2 broken rotor bars

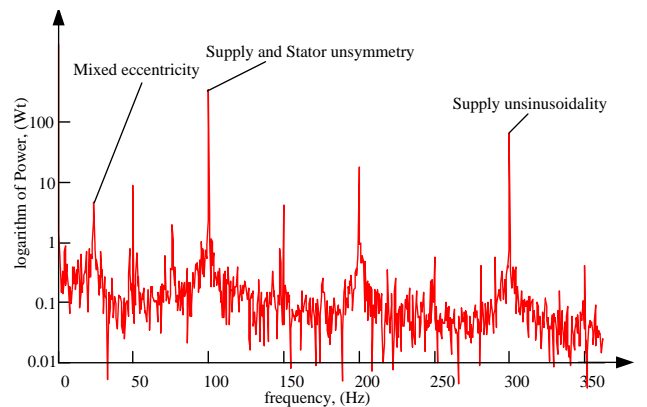
Fig. 1 Phase current (continuation)



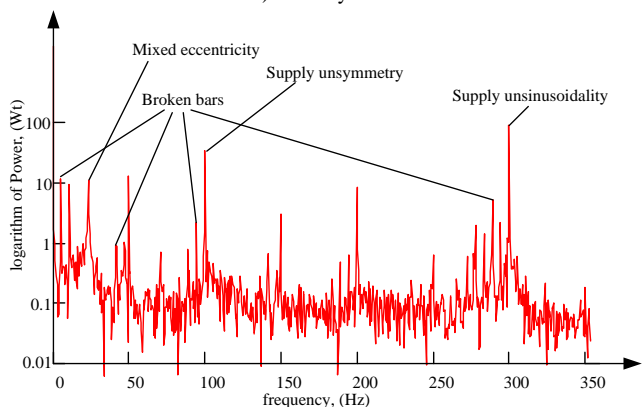
d) 2 broken rotor bars and 2.54% stator phase unsymmetry



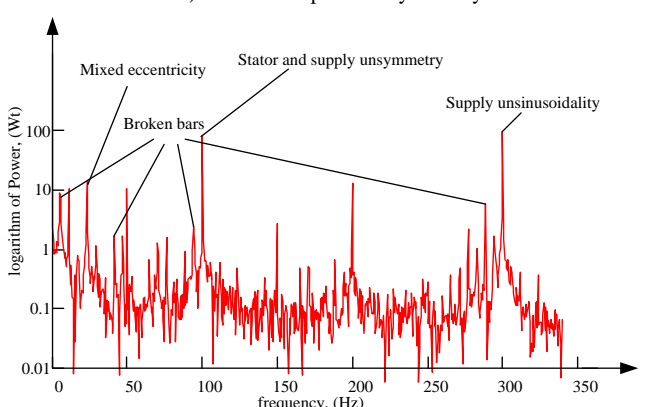
a) Healthy motor



b) 14% stator phase unsymmetry



c) 2 broken rotor bars



d) 2 broken rotor bars and 2.54% stator phase unsymmetry

Fig. 2 Total 3-phase power

V. CONCLUSION

A comparison of IM fault detection methods based on analysis of current and power signals' spectra showed, that both methods could be used for detection the most common motor damages. But current spectra analysis in some cases could lead to wrong diagnosis, because of small amplitude values of harmonics, related to damage. Power spectra analysis allows avoid such mistakes, and it could be considered as more suitable and reliable method for IM fault detection.

REFERENCES

- [1] D. Mamchur "An instantaneous power spectra analysis as a method for induction motors fault detection", Proceedings of OWD'2011, 22-25 October 2011, Wisla, pp. 407-412, ISBN 83-922242-4-0.
- [2] M. E. H. Benbouzid "A review of induction motors signature analysis as a medium for faults detection", IEEE Transactions on Industrial Electronics, vol. 47, no. 5, pp. 984-993, Oct. 2000.
- [3] C. Concar, G. Franceschini and C. Tassoni, "Differential Diagnosis Based on Multivariable Monitoring to Assess Induction Machine Rotor Conditions", IEEE Transactions on Industrial Electronics, vol. 55, no. 12, pp. 4156-4166, Dec. 2008.
- [4] A. Bellini, A. Yazidi, F. Filippetti, C. Rossi and G.-A. Capolino, "High Frequency Resolution Techniques for Rotor Fault Detection of Induction Machines", IEEE Transactions on Industrial Electronics, vol. 55, no. 12, pp. 4200-4209, Dec. 2008.
- [5] J. Cusido, L. Romeral, J.A. Ortega, J.A. Rosero and A. Garcia Espinosa, "Fault Detection in Induction Machines Using Power Spectral Density in Wavelet Decomposition", IEEE Transactions on Industrial Electronics, vol. 55, no. 2, pp. 633-643, Feb. 2008.
- [6] S.F. Legowski, A.H.M. Sadrul Ula and A.M. Trzynadlowski, "Instantaneous power as a medium for the signature analysis of induction motors", IEEE Transactions on Industrial Electronics, vol. 32, no. 4, pp. 904-909, Jul./Aug. 1996.
- [7] A.M. Trzynadlowski and E. Ritchie, "Comparative investigation of diagnostic media for induction motors: a case of rotor cage

- faults”, IEEE Transactions on Industrial Electronics, vol. 47, no. 5, pp. 1092–1099, Oct. 2000.
- [8] M. Drif and A.J.M.Cardoso, “The Use of the Instantaneous-Reactive-Power Signature Analysis for Rotor-Cage-Fault Diagnostics in Three-Phase Induction Motors”, IEEE Transactions on Industrial Electronics, vol. 56, no. 11, pp. 4606–4614, Nov. 2009.

ADC test stand with exponential excitation signal using LabVIEW

¹Ing. Jozef LIPTÁK (1st year), ²Ing. Marek GODLA (2nd year), ¹Ing. Daniel FÁBRI (ext)
Supervisor: ¹doc. Ing. Ján ŠALIGA, PhD., ²prof. Ing. Linus MICHAELI, DrSc.

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

{jozef.liptak, marek.godla}@tuke.sk, dfabri@azet.sk

Abstract—The paper deals with design and implementation of the testing system for testing analogue-to-digital converters. The system itself takes advantage of exponential excitation signal and the programming environment LabVIEW. The main part of the paper is creation of the functions library (subroutines) which can be used not only within the final program, but also within further development and testing of the measuring method if needed. Measurements were done at chosen converters using the final program. The same converters were tested by standard testing method. The results acquired by both methods are compared to validate the above mentioned implementation of the testing method.

Keywords—LabVIEW, AD convertor, testing, exponential stimulus

I. INTRODUCTION

Each AD (analog-digital) / DA (digital-analog) conversion deteriorates signal by additional errors. The basic requirement at application of AD or DA conversion is that these errors should be as small as possible.

Today's market offers a wide range of ADC's with high quality (with respect to their use). In their common applications it is not necessary to choose the "better" one. But if such a need occurs, it is necessary to use a method to do a qualitative analysis. In most cases, we get to the final INL (integral nonlinearity) graph of the converter.

In standard [7] a dynamic method for testing AD converters is described in order to determine their INL. This method is standardized and in our experiments, the results obtained with it will be considered as correct (reference). Of course, other methods of testing have been developed, each of which has its advantages and disadvantages [2]. At our department one such proposal of a test method is also developed. It is a dynamic histogram method, using an exponential bidirectional signal as a test stimulus. Initial proposals and experiments were published in [3]. This paper was focused on the generation of the exponential test signal. The final version of the proposed method is described in [4] and [1]. We implemented in our programs the mathematical procedures that were derived in [1].

Our goal was to develop software for test stand in form of library with subroutines allowing, if necessary, to be used in continued developing, or use them in other programs. The final program was tested on real ADC's in order to verify its functionality and metrological validity.

The paper is organized as follows: chapter II describes the theoretical basis of the measuring method and its mathematical apparatus. Chapter III is a description of the implementation and of the main programs. Chapter IV discusses measured results obtained from measurements with the standard method and the results obtained from the histogram method using exponential excitation signal. These are then compared.

II. THEORETICAL FUNDAMENTS

The histogram test method with exponential test signal is derived from the histogram method described in the standards [7]. It is mainly focused on the determination of INL and DNL of ADC under test. The exponential signal occurs in "nature", so it is also easier to generate a good exponential test signal using common components than to create high-quality harmonic signal. Therefore hardware implementation of the proposed method is simpler than the method described in the standards. Steps in the data processing are similar to the standard method:

- to generate a test signal
- to acquire a data record
- to build (calculate) a histogram from the record
- to determine the parameters of the test signal from the histogram
- to calculate the DNL and INL

A. Description of the method

The signal, which is used to excite ADC under test is generated by following way: the DAC generates a square wave signal with 50% duty cycle, which is then integrated by the integrating circuit that produce the final exponential signal by the charging and discharging of integrating capacitor. Output dual slope exponential signal is used as excitation signal for the ADC under test (DUT). The principle of generation is shown in Fig. 1, and the generated excitation signal in Fig. 2. The mathematical representation of the signal is given by (1)

$$u_{in}(t) = \begin{cases} (F_2 - B_f) \cdot e^{\frac{t_1-t}{\tau_f}} + B_f, & \text{for } (t_1 < t < t_2) \\ B_r - (B_r - F_1) \cdot e^{\frac{t_3-t}{\tau_r}}, & \text{for } (t_3 < t < t_4) \end{cases} \quad (1)$$

where F_1 and F_2 are the minimum and maximum voltages of the ADC (DUT) input full scale range, τ_r and τ_f are the time constants of the exponential pulse, B_f and B_r are the final values of exponential signal for $t \rightarrow \infty$ for decreasing and increasing exponential part of the signal, respectively (i.e. the maximum and minimum voltage of the rectangular signal generated in the DAC), and t_1, t_2, t_3, t_4 are times when the signal crosses the ADC input range limits F_1 or F_2 (see Fig. 2).

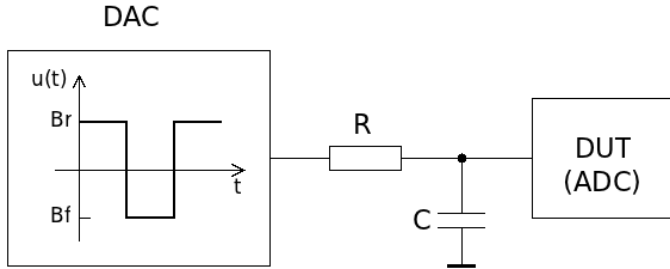


Fig. 1 Generation of the exponential test signal

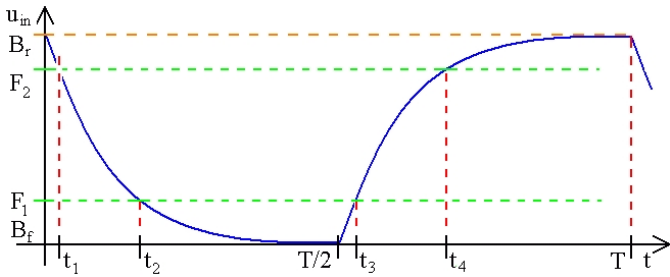


Fig. 2 Exponential excitation signal in time domain

To meet the requirements for quality of the exponential signal, we need to fulfill only simple rules:

- generated rectangular signal must have short switching times between high and low level;
- capacitor C must be of high quality (dielectricum!), so that it does not cause any additional shape distortion of the generated signal.

The best capacitor for the integration circuit was selected on the basis of experiments performed in [3]. A capacitor with a polypropylene sheet type dielectric (No. MKP X2 SH) 470nF/275VAC from the company ARCOTRONICS was used. The resistor has negligible effect on the quality of the generated signal. We used an older type of carbon film resistor 80MJIT with a resistance of 82kΩ. It is appropriate to shield the entire integration circuit to achieve a good signal to noise ratio.

The record has to contain an integer number of signal periods. Then two independent histograms are calculated from the record for the rising and for the falling part of recorded signal. The histograms have to be normalized according to the following equation

$$H_n(i) = \frac{H(i)}{\sum_{l=0}^{2^N-1} H(l)} \quad (2)$$

where $H(i)$ is a histogram calculated from the record.

B. Model of the histogram

The INL and DNL can be computed from the measured and

ideal histogram. For the ideal histogram we use a recurrent model introduced in [3]. It is defined as shown in (3), where b is the only parameter we need to estimate. This is a great advantage of this recurrent histogram model. This model has also small errors as it was shown in [5] and is easy to compute, because we need to perform just basic arithmetical operations.

$$\hat{H}_{nid}(k) = \frac{\hat{S}}{k - \frac{1}{2} - b}, \quad \hat{S} = \frac{1}{\sum_{l=1}^{2^N-2} \frac{1}{l - \frac{1}{2} - b}} \quad (3)$$

$$\hat{H}_{cnid}(k) = \sum_{i=1}^k \hat{H}_{nid}(i) = \sum_{i=1}^k \frac{\hat{S}}{i - \frac{1}{2} - b}.$$

C. Excitation signal parameters estimation

The main goal of this method is to estimate the parameters of the excitation signal from the measured histogram. Parameters that we need to estimate are the voltages in infinite time, separate for rising and falling part of the excitation signal (b_r, b_f). We achieve this by fitting measured data using the recurrent model. It is done by Newton's iteration algorithm that is used to minimize the cost function

$$\min(CF) = \min(\phi(b_r, b_f)) = \min \left(\sum_{i=1}^{2^N-2} (INL(i, b_r) - INL(i, b_f))^2 \right) \quad (4)$$

where $\phi(b_r, b_f)$ is the cost function, and b_r and b_f are the unknown parameters. The INL and DNL are computed as

$$DNL(i, b_r) = \frac{H_{nr}(i) - H_{nid}(i, b_r)}{H_{nid}(i, b_r)} = \frac{H_{nr}(i)}{H_{nid}(i, b_r)} - 1,$$

$$DNL(i, b_f) = \frac{H_{nf}(i) - H_{nfd}(i, b_f)}{H_{nfd}(i, b_f)} = \frac{H_{nf}(i)}{H_{nfd}(i, b_f)} - 1, \quad (5)$$

$$INL(i, b_r) = INL(i-1, b_r) + DNL(i, b_r),$$

$$INL(i, b_f) = INL(i-1, b_f) + DNL(i, b_f).$$

where $H_{nr}(i)$ and $H_{nf}(i)$ are the measured normalized histograms, and b_r, b_f are the normalized parameters B_r, B_f . Because we do not take into account the actual voltage at the ADC input and record only codes from the ADC output (raw data) and also the calculations are made in code domain, we can assume that $b = B$.

The local minimum of the function (4) is given by

$$\frac{\partial \phi(b_r, b_f)}{\partial b_r} = 0, \quad \frac{\partial \phi(b_r, b_f)}{\partial b_f} = 0. \quad (6)$$

The Newton iteration process for this system of equations is given by (7). To simplify the calculations we don't use the exact values of the partial derivations needed in (7), but just their approximations with the corresponding Taylor series, when we suppose that $h_r \approx h_f \approx h$. They are not described here due to space limitations for this paper. You can find them in [1]. The STOP criterion is, according to [4], derived from assumption:

$$\varepsilon \approx h^3, h \approx \varepsilon^{1/3} \quad (8)$$

where ε is the machine epsilon (it represents the round-off error for a floating-point number with a given precision) and h is the increment. Finally the new parameters b_r and b_f are computed as shown in (9).

$$\begin{aligned} \frac{\partial^2 \phi(b_r, b_f)}{\partial b_r \partial b_f} h_r + \frac{\partial^2 \phi(b_r, b_f)}{\partial b_r \partial b_f} h_f &= - \frac{\partial \phi(b_r, b_f)}{\partial b_r}, \\ \frac{\partial^2 \phi(b_r, b_f)}{\partial b_r \partial b_f} h_r + \frac{\partial^2 \phi(b_r, b_f)}{\partial b_r \partial b_f} h_f &= - \frac{\partial \phi(b_r, b_f)}{\partial b_f}, \end{aligned} \quad (7)$$

$$b_r^{[n]} = b_r^{[n-1]} + h_r, \quad b_f^{[n]} = b_f^{[n-1]} + h_f.$$

$$\begin{aligned} b_r^{[n+1]} &= b_r^{[n]} - \frac{\frac{\partial \phi}{\partial b_r} \cdot \frac{\partial^2 \phi}{\partial b_r^2} - \frac{\partial \phi}{\partial b_f} \cdot \frac{\partial^2 \phi}{\partial b_r \partial b_f}}{\Delta}, \\ b_f^{[n+1]} &= b_f^{[n]} - \frac{\frac{\partial \phi}{\partial b_f} \cdot \frac{\partial^2 \phi}{\partial b_r^2} - \frac{\partial \phi}{\partial b_r} \cdot \frac{\partial^2 \phi}{\partial b_r \partial b_f}}{\Delta}, \end{aligned} \quad (9)$$

$$\Delta = \begin{vmatrix} \frac{\partial^2 \phi}{\partial b_r^2} & \frac{\partial^2 \phi}{\partial b_r \partial b_f} \\ \frac{\partial^2 \phi}{\partial b_r \partial b_f} & \frac{\partial^2 \phi}{\partial b_f^2} \end{vmatrix} = \frac{\partial^2 \phi}{\partial b_r^2} \cdot \frac{\partial^2 \phi}{\partial b_f^2} - \left(\frac{\partial^2 \phi}{\partial b_r \partial b_f} \right)^2.$$

The STOP criterion of the iteration process is $|b_i^{[n+1]} - b_i^{[n]}| \leq \varepsilon_i$, where ε_i in (8) is the chosen residual uncertainty. The final $b_r^{[n+1]}, b_f^{[n+1]}$ are used to calculate $DNL_r(k)$, $DNL_f(k)$, $INL_r(k)$ a $INL_f(k)$ of the tested ADC using (5). The final INL and DNL (i. e. main result) is calculated as

$$\begin{aligned} DNL(k) &= \frac{DNL_r(k) + DNL_f(k)}{2}, \\ INL(k) &= \frac{INL_r(k) + INL_f(k)}{2}. \end{aligned} \quad (10)$$

III. IMPLEMENTATION OF THE METHOD

The generator was built on multifunction DAQ card PCI-6289 with the termination unit BNC-2120 by NI. The integration circuit was shielded with a metal box. All connections were realized by coaxial cables.

Two separate software routines VI (Virtual Instruments, programs in LabVIEW, see [6]) were developed. One VI collected data and calculated histograms. The other VI performed analysis and final calculations of the DNL and INL. The programs communicated using two files contain separated histograms for the rising and falling exponential signal slopes.

A. Program MEASURE.VI

This program enables setting test conditions, such as sample rate, Hi and Lo levels of the generated square wave, etc. The program shows in real time the collected histograms as they grow and automatically stops the data collecting after achieving the required number of samples (or after the user

presses the STOP button).

B. Program ANALYZE.VI

This is the main program for this method. It performs fitting of the model histogram and calculates the final INL. The main parts of it are the output graph windows. It can also save the results in a specified file.

The complete documentation of these programs, including screenshots of the GUIs and the source code is in [1].

IV. EXPERIMENTAL RESULTS

To test the new implementation in LabVIEW and to verify its metrological performance we have performed a number of tests on real ADCs. The results have been compared with reference results acquired by the standardized histogram test method using the precise generator STANFORD DS360 and software developed in [3]. Because of space limitations we show here only test results of the following tested ADCs: USB6009 (14bit), ADC0804LCN (8bit). Test conditions are briefly described in the following figures captions.

A. USB6009

It is an external measurement card from NI, with a USB connection. The Fig. 5 shows the difference between the reference result, and the result obtained by the new method.

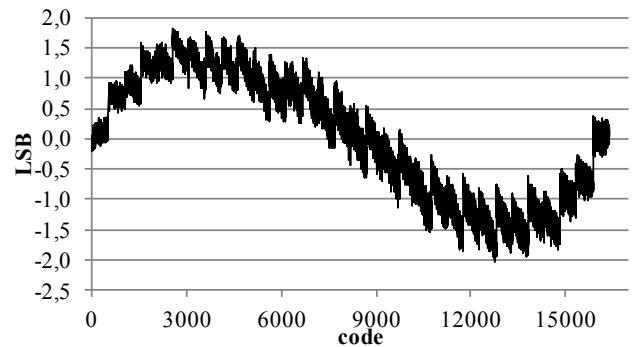


Fig. 3 INL of USB6009 using standardized method (frequency of excitation signal: 111.111Hz, resolution: 14bit, input range: $\pm 1V$, sampling rate: 48 000Ss, # of samples in record: 8×10^6)

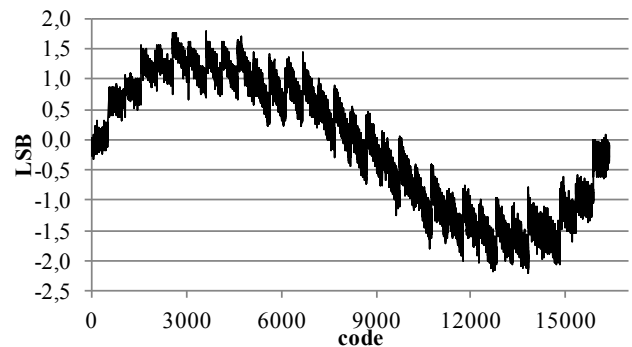


Fig. 4 INL of USB6009 using the discussed method (frequency of excitation signal: 23.5Hz, resolution: 14bit, input range: $\pm 1V$, sampling rate: 48 000Ss, # of samples in record: 8×10^6)

The theoretical maximal difference shown in Fig. 5 according to the number of samples in record should be 0.6LSB (according to [3]). The requirement is accomplished, what is a positive proof of the validity of the discussed method.

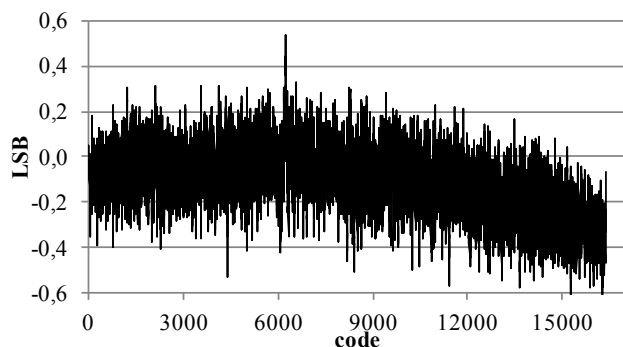


Fig. 5 Difference of INLs in Fig. 4 and Fig. 3

B. ADC0804LCN

It is one chip ADC. The connection to the PC and whole setup was made according to [8]. Fig. 8 shows the difference between the reference result and the result obtained by the new method.

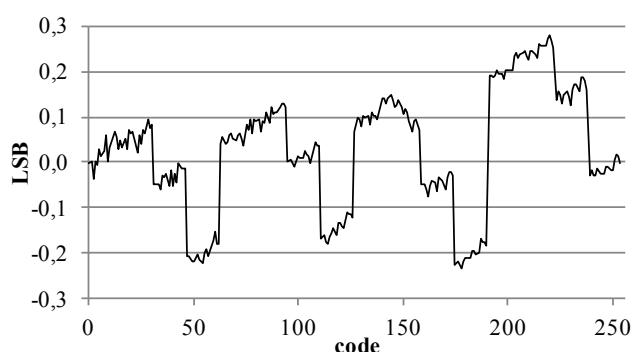
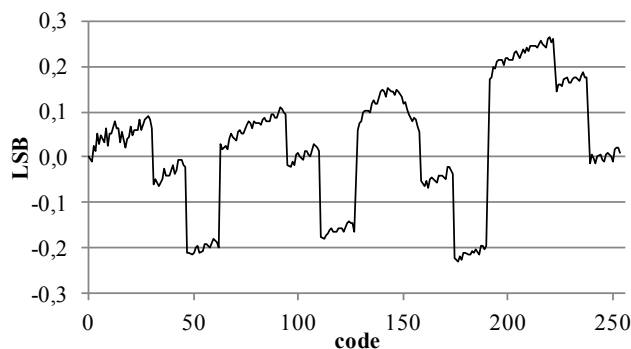
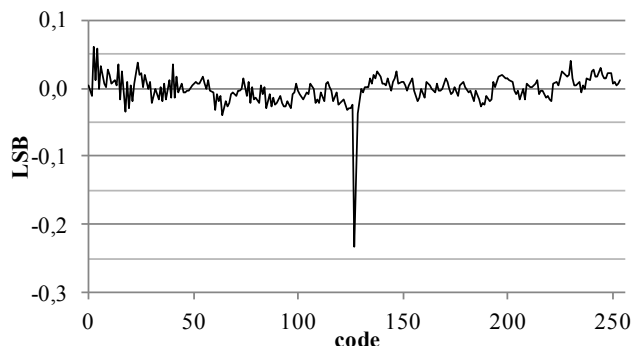
Fig. 6 INL of ADC0804LCN using standardized method (frequency of excitation signal: 111.111Hz, resolution: 8bit, input range: 0÷5V, sampling rate: 8 110Ss, # of samples in record: 2×10^6)Fig. 7 INL of ADC0804LCN using the discussed method (frequency of excitation signal: 16.5Hz, resolution: 8bit, input range: 0÷5V, sampling rate: 8 110Ss, # of samples in record: 2×10^6)

Fig. 8 Difference of INLs in Fig. 7 and Fig. 6

This ADC had the biggest difference in the acquired INLs in the code 127. This difference exceeds the theoretical measurement inaccuracy that should be max. 0.12LSB. The code 127 of this ADC appears to have a build-in/construction error because the same effect was observed also in earlier experiments, including static tests.

The differences of INLs shown in Fig. 5 and Fig. 8 have a noise-like character, what proofs, that there is no systematic error interleaved into the testing, and the difference is caused just due statistic uncertainty of both test methods.

V. CONCLUSION

This method has great potential, because using the methods described in the standards it is already unrealistic to test present high-resolution ADC. The development of new testing methods is therefore necessary.

In the future we would like to modify the GUI of the main program. Another desired modification of the program for obtaining histograms is that it should automatically recognize the format of the raw data obtained from the tested ADC. The comfort of the operating personnel would improve, if the program ANALYZE.VI had an algorithm that guaranteed even without entering the initial B_f , B_r , that the program returned a result. Due to problems with convergence of the proposed minimization in the test method it is sometimes necessary to repeatedly change the initial value of B_f , B_r to get the right results.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] LIPTÁK, Jozef : Testovanie AČ prevodníkov. Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2011. 57 s.
- [2] SAKMÁR, Michal : Testovanie analógovo-číslicových prevodníkov. Diplomová práca. Košice: Technická univerzita v Košiciach, FAKULTA ELEKTROTECHNIKY A INFORMATIKY, 2006. 51 s.
- [3] SAKMÁR, Michal : Využitie netradičných signálov pre testovanie analógovo-číslicových prevodníkov. Dizertačná práca. Košice: Technická univerzita v Košiciach, FAKULTA ELEKTROTECHNIKY A INFORMATIKY, 2009. 103 s.
- [4] ŠALIGA, Ján et al. : Processing of bidirectional exponential stimulus in ADC testing. In: Measurement, Kidlington: Elsevier, 2010. s. 1061 - 1068.
- [5] MICHAELI, Linus et al. : New approximation models for processing exponential stimulus histogram test of ADC, In: 16th IMEKO TC-4 Symposium, Florence, Italy, September 22–24, 2008.
- [6] National Instruments Corporation : LabVIEW™ Help. [LabVIEW program help]. Actualized 6-2011, under number 371361H-01 [cit 2012-02-19].
- [7] IEEE Std. 1057 IEEE Standard for Terminology and Test Methods for Analog-to-Digital Converters, 2007.
- [8] National Semiconductor : ADC0801/ADC0802/ADC0803/ADC0804/ADC0805 8-Bit μ P Compatible A/D Converters. [integrated circuit datasheet]. Actualized 11-1999, under number DS005671 [cit 2011-03-10].

Application binary state arithmetic encoding in JPEG

¹Ondrej KOVÁČ (1st year), ²Kornel RUMAN (1st year), ³Ján VALISKA (1st year)
Supervisor: ⁴Ján MIHALÍK

^{1,3,4}Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic
²Dept. of Technologies in Electronics, KTE TU of Košice, Slovak Republic

¹ondrej.kovac@tuke.sk, ²kornel.ruman@tuke.sk, ³jan.valiska@tuke.sk, ⁴mihalik.jan@tuke.sk

Abstract— This paper shortly describes basic image JPEG format. This paper also describes issue of decomposition of DCT image spectral coefficients to its bit levels and their encoding using binary state arithmetic code. In the conclusion, the advantage for replacement of conventionally used Huffman code by binary state arithmetic encoding of the picture's bit levels DCT spectral coefficients image is contemplated.

Keywords— JPEG, DCT compression, bit plane and arithmetic encoding.

I. INTRODUCTION

The transmission speed of connection has been increasing, but given the fact that the amount of data transferred also increases, it is compulsory to enhance the methods of compression applied to the multimedia data.

This article is focused on an algorithm of source encoding in JPEG. In JPEG, the Huffman encoding is being used. We will try to replace this algorithm using the binary state arithmetic encoding.

II. JPEG WITH DISSIPATION ENCODING

Block diagram of image encoder is on figure 2.1. There can be colorful or grey scale image on the input. This structure corresponds with intern-image transformation encoding with DCT block size 8x8 pixels and with entropic encoding.[1]

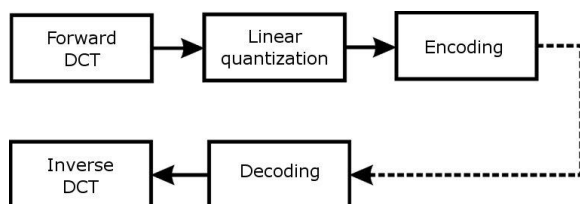


Fig. 2.1. Block diagram image codec JPEG with dissipation encoding.

At the beginning, the image is divided to 8x8 block pixels and after this, DCT is applied. All spectral coefficients (SC) are quantized with different or equal quantization step, which changes, depending on statistical properties or application. Quantized SC are scanned in a diagonal order called with quantization step. [2] Quantized SC are scanned in diagonal order called Zig-zag scanning. Zig-zag is showed on figure 2.2. By using this scanning, we achieve that SC with high

energy end small frequency will be scanned before SC with high frequency but small energy do so. SC with zero energy are detected and counted. This allows making couples “number of zeros and non-zero values of SC”. These couples are entropic encoded by Huffman code. [3]

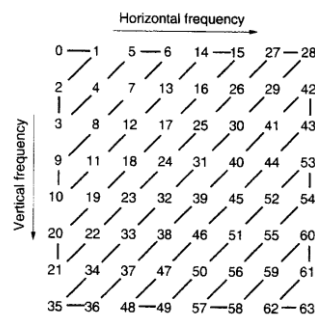


Fig. 2.2 Zig-zag scanning

Each one dimensionally represented block must be ended with special couple [0][0] called EOB (end of block).

A. Used quantizers

By quantization in space of SC, each of SC is quantized, using its own quantizer. This is showed on figure 2.3.

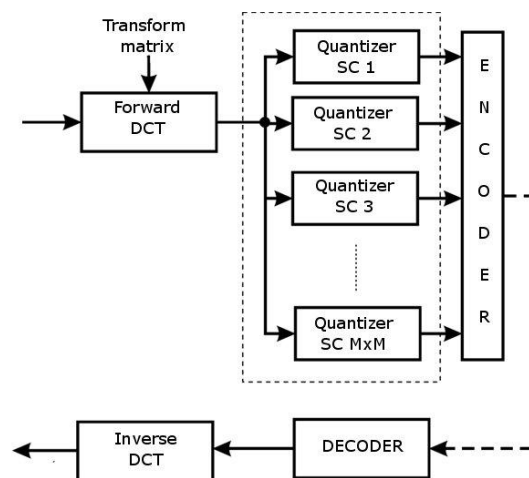


Fig. 2.3 Block diagram of transformation encoding system

In this letter we will use two types of quantizers. We will use one for DC spectral coefficients (DSC) and one for AC spectral coefficients (ASC).

1) ASC quantization

For quantization of ASC, the used algorithm is described by the equation 2.1

$$q_{ij} = 8Y_{ij} // MK\Delta_{ij} \quad (2.1)$$

Where q_{ij} represents an index of quantization and takes its values on interval $\langle -255, 255 \rangle$. Y_{ij} is a value of original spectral coefficient and MK is a scaling factor. Δ_{ij} is quantization step. Operation $//$ means dividing with rounding on nearest integer.

Results of this operation are quantization indexes. [2]

Reverse operation is described by equation 2.2.

$$\tilde{Y}_{ij} = q_{ij}MK\Delta_{ij} / 8 \quad (2.2)$$

2) DSC quantization

For quantization DSC, algorithm represented by equation 2.3, is used.

$$q_{00} = Y_{00} // 8 \quad (2.3)$$

Usually quantization of DSC is performed with fix set step. The obtained quantization level is showed in equation 2.4.

$$\tilde{Y}_{00} = 8q_{00} \quad (2.4)$$

III. IMAGE DECOMPOSITION TO BIT PLANE

In general, each image can be expressed like two dimension rasters of pixels and each pixel can be described by N bits as follows:

$$\text{pixel} = p_0 2^0 + p_1 2^1 + p_2 2^2 + \dots + p_{N-1} 2^{N-1}$$

Individual bits P_i correspond to the respective weights 2^i for $i=0, 1, \dots, N-1$. If we pick up only P_0 from in a such way expressed pixel from entire polytional image, we get its bit plane of zero weight (BP_0). Using the same process all other bit-planes until BP_{N-1} are created.

In a transformed space, there are not only spectral coefficients with positive value. Because of this fact, there are two various ways for decomposition [2].

- *Decomposition of the real bit-plane (RBP)*
- *Decompositions of the absolute bit-plane and plane of sign (ABP)*

We not decompose image of spectral coefficients, but image of quantization indexes in range $\langle -255; 255 \rangle$.

A. Decomposition of the real bit-plane (RBP)

In this decomposition, we get bit-planes $-BP_{N-1} \dots -BP_0 \dots BP_0 \dots BP_{N-1}$. Figure 3.1. shows the block diagram of RBP decomposer.

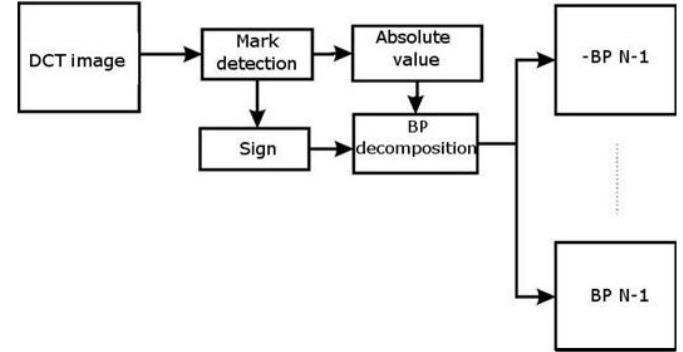


Fig. 3.1 Block diagram of RBP decomposer

Each of spectral coefficients of transformed image is tested on a sign and the result of this is a mark defining either positive or negative bit plane. After this, the image is transformed into absolute value and decomposed with identical method like decomposition of image in Gaussian space.

B. Decompositions of the absolute bit-plane and plane of sign (ABP)

In this decomposition, we get bit-planes $BP_0, BP_1 \dots BP_{N-1}$ and BP_s . Block diagram of ABP decomposer is showed on figure 3.2.

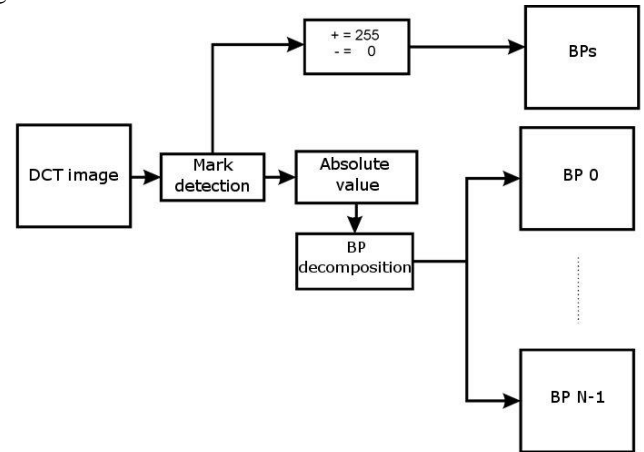


Fig. 3.2 Block diagram of ABP decomposer

Each of spectral coefficients of transformed image is tested on the sign and the result shows value of pixel in bit-plane BP_s . This value is 255, or white for positive and 0, or black for negative mark. After this, the image is transformed on absolute value and decomposed with identical method like in previous case.

IV. ARITHMETIC ENCODING

A. Arithmetic code without using multiplication

1) Encoding

Procedure of the arithmetic encoding (AE) without using multiplications is described with equations 4.1 – 4.4. [4]

$$A(sk) = A(s) \cdot 2^{-Q} \quad (4.1)$$

$$A(sm) = < A(s) - A(sk) > \quad (4.2)$$

$$C(sm) = C(s) \quad (4.3)$$

$$C(sk) = C(s) + A(sm) \quad (4.4)$$

Where $C(s)$ represents binary variable of sequence S with 0 initial value before encoding first symbol, and expresses low limit of probabilistic subinterval. $C(sm)$ is binary variable of sequence S for the more probable symbol and $C(sk)$ for less probable symbol of sequence S .

Size of probabilistic subinterval is extended by variable $A(s)$, which initial value is 0.111....11. Variables $A(sk)$ and $A(sm)$ are binary variables of size of probabilistic subinterval for more and less probable symbol.

This method is based on approximation of probability of less probable symbol by value of 2^{-Q} . Variable $C(s)$ is obtained by gradual encoding of symbols of binary variable S . When all symbols of binary variable S are encoded we get two binary variables, $C(s)$ and $A(s)$. Resulting arithmetic code is a binary value from interval $<C(s), C(s) + A(s)>$. This value is made, so it contained the smallest possible amount of valid binary numbers.

2) Decoding

Procedure of decoding is described by equations 4.5 – 4.8.

$$A(sk) = A(s) \cdot 2^{-Q} \quad (4.5)$$

$$A(sm) = < A(s) - A(sk) > \quad (4.6)$$

$$C(s) < A(sm) \Rightarrow y = m; C(sy) = C(s) \quad (4.7)$$

$$C(s) \geq A(sm) \Rightarrow y = k; C(sy) = C(s) - A(sm) \quad (4.8)$$

Input encoded sequence is decoded with gradual recursion of equations 4.5 – 4.8. Decoder must have information about length of encoded word. This information could be broadcasted in transmission channel separated from encoded sequence, or the length can be given in advance.

B. Binary state arithmetic encoding (BSAC) of binary images

In arithmetic encoding without using multiplications, it is necessary to know value of Q from expression 2^{-Q} .

This value must be given for each symbol of sequence. This information is contained in the model [5]. Model is practically an estimator estimating probability of actual symbol occurrence. JBIG algorithm is used For binary images. Ten points template of JBIG algorithm is on figure 4.1.

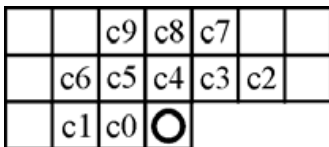


Figure. 4.1. Ten points template of JBIG algorithm.

State $S(y)$ of actual point (indicated by circles) is estimated by equation 4.9.

$$S(y) = \sum_{i=0}^{n-1} ci \cdot 2^i \quad (4.9)$$

Where n is a number of elements in template and y is an actual point. After testing the whole image, it is evaluated how many times a particular state occurred and which value actual point had. After that, the most frequent occurred value is determined. Finally, conditional probability will be calculated (eq. 4.10.) of point with value “1” or “0” in individual states.

$$p(c = v | S_k) = p(S_{v,k}) = \frac{n_{v,k}}{n_k} \quad (4.10)$$

S_k is value given by equation 4.9 and mean value of actual pixel. $S_{v,k}$ is state of S_k with value of pixel equal to v . n_k is a number of states $S_{v,k}$. The final model of image is formed by the resulting probabilities. Conditional probability $S_{v,k}$ are approximated by value 2^{-Q} .

V. EXPERIMENTAL RESULTS

Experimental images LENA and BABOON fig. 5.1 and 5.2, respectively, were transformed by DCT in to images of spectral coefficients fig. 5.3 and 5.4. After that, they were quantized with two scales. Lower value of scale is set so that the IDCT (inverse discrete cosine transform) subjective quality would be approximately the same as the original images. High level is set to maximum possible value therefore 32. On fig. 5.5 and 5.6 images after quantization and IDCT are showed.



Fig. 5.1 Image Lena

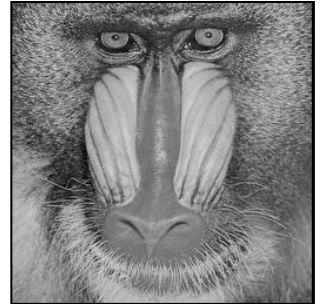


Fig. 5.2 Image Baboon

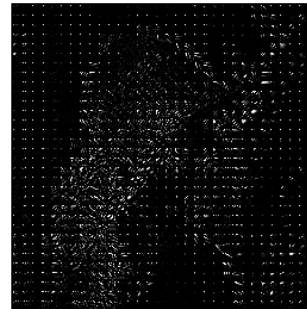


Fig. 5.3 DCT image Lena

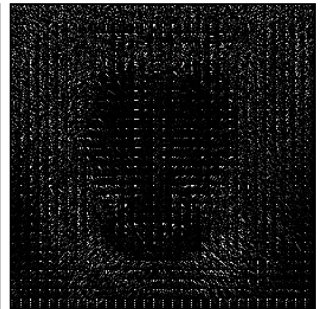


Fig. 5.4 DCT image Baboon



a)



b)

Fig. 5.5 IDCT image Lena a) MK = 15 b) MK = 32

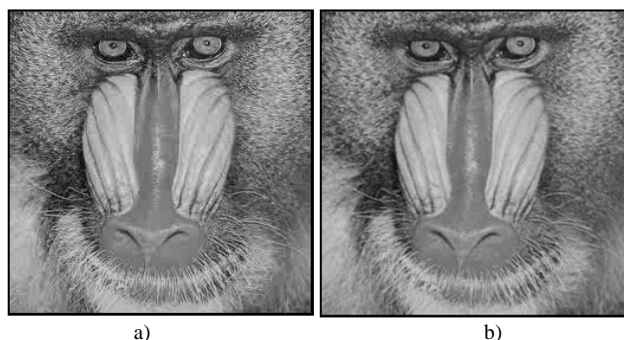


Fig. 5.6 IDCT image Baboon a) MK = 17 b) MK = 32

Table 1 shows middle length of code words of images encoded by BSAC and HC and also differences between middle lengths of code words.

TABLE I

DIFFERENCE BETWEEN MIDDLE LENGTH OF CODE WORD OF HC AND BSAC

Image		Middle length of code word		Difference between HC and BSAC		
Image	Scale	BSAC bit/sc	HC bit/sc	Reduction of code word	Extension of code word	
RBP	LENA	15	1,7679	1,8800	0,1121	-
		32	1,1489	1,4680	0,3192	-
	BABBBON	17	2,3959	2,2130	-	0,1829
		32	1,6004	1,6100	0,0096	-
ABP	LENA	15	1,7982	1,8800	0,0818	-
		32	1,2082	1,4680	0,2598	-
	BABBBON	17	2,4980	2,2130	-	0,2850
		32	1,7526	1,6100	-	0,1426

From the table, it can be concluded that maximum saving is achieved by using BSAC applied on RBP for images with low frequency content therefore for an image with lot of monotone areas.

For an image with high frequency content, it is better to use conventional encoding, therefore Huffman encoding.

BSAC implicated on the image decomposed to ABP is possible, but it gets lower compression like application on RBP. From the table, it is also evident that with the increase of MK there is increasing compression in both encoding methods compression. This increase is stronger in BSAC than in HC. This is illustrated in table 2.

TABLE II

INCREASING OF COMPRESSION FOR BSAC AND HC WITH INCREASING OF SCALE OF QUANTIZATION

		Middle length of code word		Increasing of compression			
		BSAC		HC		BSAC	
Image	Scale	RBP	ABP	bit/sc		ABP	RBP
LENA	15	1,7679	1,7982	1,880			
	32	1,1489	1,2082	1,468	0,412	0,590	0,6191
BABOON	17	2,3959	2,4980	2,213			
	32	1,6004	1,7526	1,610	0,603	0,7454	0,7955

VI. CONCLUSION

In the article, we discussed the modification of standard JPEG using arithmetic binary-state encoding applied on bit-planes of DCT image instead common used Huffman code. With experiments we arrive at the results which suggest that increased degree of compression for images with low frequency content. Conversely, this encoding doesn't achieve satisfactory results for images with high frequency content. It was found out that it is better to encode the real bit-plane of DCT image than absolute bit-plane. We achieved the best results using BSAC on image quantized with high MK. It can be inferred, that this modification of JPEG can be used, but it is necessary to add other algorithms which detect level of details of image and decides what kind of encoding to use. Another disadvantage is the necessary decomposition of the image to bit-plane and encoding of these using more difficult arithmetic code and a need to determine actual state and conditional probability. These disadvantages cause increased demands on hardware performance.

REFERENCES

- [1] Mihalík, J.: Image Coding in Videocommunications. Mercury-Smékal, Košice, 2001. ISBN 80-89061-47-8 (244 pages). (In Slovak)
- [2] Kováč O.: Transformačné kódovanie obrazov. Bakalárska práca FEI TU Košice, 2009.
- [3] Mihalík, J.-Štefanišin, R.: Entropické kódovanie v štandardných videokodekoch. Acta Electrotechnica et informatica, Vol.5, No.2, 2005, s.43-50, ISSN 1335-8243.
- [4] Gladišová, I.-Mihalík, J.-Zavacký, J.: Bezstratová kompresia obrazu pomocou stavového binárneho aritmetického kódovania jeho bitových rovín. Slaboproudý obzor, roč. 62, č.1-2, 2006, str.18-21, ISSN 0037-668X.
- [5] Gladišová, I. - Mihalík, J.: Stavové aritmetické kódovanie binárnych obrazov. Acta Electrotechnica et Informatica, Vol.3, No.3, Košice, 2003, p.36-44, ISSN 1335-8243.

Availability of Hybrid FSO/RF System

¹Pavol MIŠENČÍK (3st year), ²Tomáš HARASTHY (2nd year)
Supervisor: ³Ján TURÁN

^{1,2,3}Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

¹pavol.misencik@tuke.sk, ²tomas.harasthy@tuke.sk, ³jan.turan@tuke.sk

Abstract – Free Space Optics technology with big bandwidth range has become wireless alternative to optical fiber network. The main advantages of FSO in opposition to fiber are its fast installation and visible savings in expenses. Disadvantage of FSO is that the received power is variable and hard to predict. It is due to the influence of the weather on the transmitted optical beam. This paper deals with the availability and properties of hybrid FSO/RF line that provides high accessibility and continuous communication.

Keywords—availability of FSO system, hybrid FSO/RF system, FSO system, FSO link

I. INTRODUCTION

FSO communication known as well as wireless optical communication is one of less developed broadband technologies despite some long-lasting problems that are described in [1]. Point-to-point FSO transmission system is able to achieve data speed comparable to data speed of fiber optic communication systems without the rise of cost and time for the installation in comparison with fiber optic. General use of FSO communication systems is hindered though due to its reliability problems. Reliability of FSO communication line markedly lowers in presence of fumes and fog. Due to that the demanded availability 99,999% is not achieved. Hence, in the recent time there is a certain increase in taking into account the idea of hybrid FSO/RF communication system being the solution of availability problem. Low speed data transmission serves as backup connection to secure data communication in time when FSO line is not running. FSO and radio-frequency (RF) lines provide various properties. For instance, FSO line suffers in consequence of atmospheric turbulence causing variance of atmospheric refraction indexes caused by fog and the like [2]. But these conditions have no effect on RF line. FSO line provides huge license-free zones while RF line spectrum is generally licensed and restricted. RF line suffers in consequence of impact of rainfalls while atmospheric loss of FSO line is significantly lower. Hybrid channel is able to use advantages of both FSO and RF lines. FSO line backed up by RF line uses strongpoint properties of each system to overcome their weaknesses presented in paper [2, 3]. In this way the hybrid FSO/RF system can achieve higher line availability then is achieved by either FSO line or RF line alone, and that is because RF part of hybrid FSO/RF is able to reduce the influence of thick fog and FSO line is able to reduce the influence of rainfalls. Fog is the most important

negative factor of FSO line and rainfall is the most influential negative factor of RF line.

II. APPLICATION AND MAIN CHARACTERISTIC OF FSO/RF SYSTEM

Hybrid FSO/RF transmitting line is dealt with by reason of back upping primary FSO transmission line via RF line which can stop operate because of negative influences, particularly fog. This hybrid system consisting of two lines works in parallel and RF line substitutes FSO primary line when it stops operate. FSO line in hybrid FSO/RF makes LightPointe system FlightStrata 155E, which we have in our laboratory in KEMT FEI TUKE. RF line consists of 60GHz line. Because FSO works with data speed up to 1Gb/sec. we use RF line with mm waves to ensure the most advantages of FSO line. More advantages of RF 60GHz line are to be presented in the following chapter.

A. FSO characteristics

FSO optical communication system is used the most for OOK/DD (*On-Of Keying/Direct Detection*) modulation. Coherent methods of coding are investigated because of the aspect of better economic use. The most known from them is PSK (*Phase Shift Keying*) differential coding with heterodyne detection.

As was mentioned above, the worst negative effects on FSO line are caused by fog which can cause it to come to a halt. The intensity is assigned by visibility distance. Some of examples of atmospheric loss and visibility for FSO communication system with wave length 850nm will be considered now. Light fog allows visibility approximately 10km and causes atmospheric loss of FSO line from 1 to 4dB/km. Heavy fog allows visibility less than 0.5km and causes huge atmospheric loss of FSO line which is bigger than 40dB/km and can raise up to 300dB/km. Atmospheric loss caused by fog depends on wave length of transmitted signal. It is lower on longer wave lengths.

Atmospheric loss caused by heavy rainfall is not minor and depends on the intensity of rainfalls. Heavy rain falls, from 90 to 100mm/h, occurring in mild climate cause atmospheric loss about 20dB/km.

B. RF characteristics

RF systems use simple modulations similarly as FSO, but from different reason. Wave technology is not very developed in these times. Transmitters with very low phase signal noise and receivers with very high linearity do not exist (at least they are not obtainable at low costs). On the other hand, this

frequency is seldom used and antennas have narrow beams. Consequent use of higher level of modulation (e.g. Quadrature Amplitude Modulation (QAM) with $M > 4$) is neither convenient, nor needed. OOK, BOSK, BFSK or QPSK are applied for systems with transmission speed 1Gbit/s currently. (Important change in transmission speed (e.g. up to 10Gbit/s) would certainly need higher level of modulation mainly because of the width of frequency wave length). Fog causes certain losses in RF transmission channel, but these losses are in comparison with losses in FSO communication channel negligible. These losses caused by fog depend on volume of rain drops in the air and on the temperature. But in comparison with FSO, main cause of connection failures in mm waves is heavy rain. Atmospheric loss of RF line might only be 25-35dB/km during 90mm/h rainfalls and it is only weakly dependable on the frequency.

From the point of view of availability, the most suitable backup lines are those which operate with 10GHz frequency (e.g. [4]) where the influence of fog and rain on the transmitted signal is small. This solution is but hardly obtainable from various reasons. The transmission speed of back up line would have to be at least 100Mbit/s and later on possibly 300-1000Mbit/s what is over the actual technical limits. One of the great FSO advantages is its license-free wave-length, but at the same time this advantage is hardly obtainable for the backup RF lines with 10GHz frequency. (This does not apply for 2.4 and 5GHz ISM/UNI wave-lengths but the overload in these wave-lengths is enormous.)

Backup made by mm waves might be proposed and being solved in different frequencies. Our model proposes RF line which is composed of 60GHz line for its specific characteristics which are closely addressed in the following chapter. 60GHz line is in fact for free (license-free), it is without interference inside the system and the wave-length width is sufficiently high even for Gb/sec transmissions. Nevertheless, this wave-length frequency can be used only for short-distance connection, typically up to 1km.

C. Hybrid FSO/RF system

Hybrid FSO/RF system depicted on the Fig.1 is composed of FSO LightPointe E155 system. Technical parameters of this system are presented in Tab.1. The availability of LightPointe E155 system in Košice is decided upon via these parameters and the data considering visibility provided by SHMU (SLOVAK HYDROMETEOROLOGICAL INSTITUTE).

Backup RF line is composed of 60GHz line and it is in operation only when the primary FSO line fails. 60GHz line is proposed due to its advantageous characteristics described in [5], these are:

- Unlicensed operation – no need to spend significant time and money to obtain a license from FCC.
- Highly secure and not interfering service – resulting from short-distance transmission length caused by absorption of O_2 and thin width of emitted beam.
- Possibility of multiple usage of one frequency – this is an advantage in communication between several customers in the same geographical region.
- Possible optical speed of data transmission – available is 7GHz bandwidth in comparison with less than 0.3GHz in another not licensed frequency spectrum.

- Mature technology – long history of this spectrum being used for secure communications.
- Carrier-class communication links enabled – 60 GHz links can be engineered to deliver.
- “Five nines” of availability if desired.

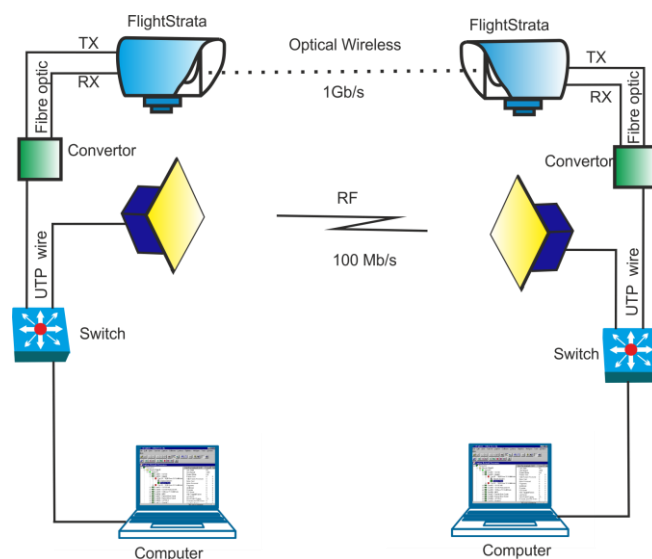


Fig. 1. Hybrid FSO/RF system

III. AVAILABILITY OF HYBRID FSO/RF LINE

We will by means of FSO System Simulator (FSO SystSim) designed and implemented in KEMT FEI TUKE when and under what visibility will FSO line composed of LightPointe FlightStrata 155E system be able to transmit. Needed input

TABLE I
FLIGHTSTRATA 155 SYSTEM SPECIFICATION

FlightStrata 155E and 155EW	
Description	Four TX, Four RX System
Dimensions	11.8 x 11.8 x 25 inches 30.0 x 30.0 x 64.0 cm
Weight	29.7 lbs / 13.5 kg
Linkhead input voltage	12 – 16 VDC
Power Supply operating voltage	115 / 230 VAC (50/60 Hz) or ± 48 VDC
Power consumption	Max. 20 W
Operating temperature	-13o to +140o F (-25o to +60o C)
Relative humidity	Up to 95% (non-condensing)
Bandwidth	1.5 to 155 Mbps
Optical transmitter	VCSEL
Output wavelength	850 nm
Beam divergence	FSA 155E 2.0 mr FSA 155EW 2.8 mr
Laser Output Power	FSA 155E, 155EW 24 mW
Receiver Sensitivity	FSA 155E, 155EW -45 dBm
Optical Interface	SC connector standard polish
Receiver Type	Si APD

data for this system are presented in Tab.1. The main parameter of hybrid FSO/RF line is the length of transmission line. This means that how far apart the transmitter and the receiver are. Our hybrid line will operate in 1km distance and for that distance will be attributed availability of hybrid line in Košice. As was mentioned before, through FSO SystSim is defined the level of visibility when FSO line will not be able to operate. On the Fig. 2 is the output from FSO SystSim.

Visibility magnitude 515m is the borderline value below which FSO line stops working.

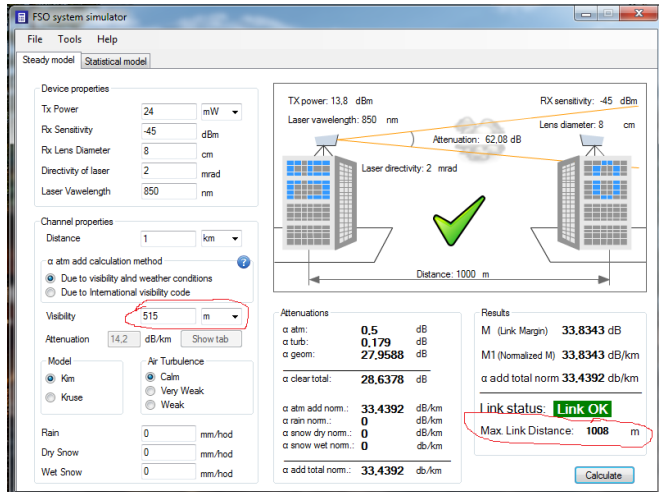


Fig. 2. FSO SystSim, Link OK

In collaboration with we obtained data of measured visibility in the area of Košice, to be precise data were from Košice airport. SHMU provided us with the measurements during whole year 2008. From these data we need to filter those which fall below 510m.

These data are depicted in the graph on the Fig. 3.

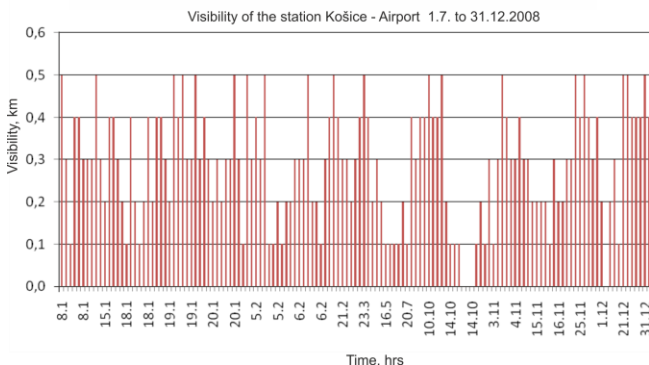


Fig. 3. Visibility in Košice <510m

On Fig. 3 are depicted only filtered measurements of visibility which value is less than 510m. Total time during which these visible distances occurred are evaluated as the sum of partial values. All the values provided by SHMU were measured hourly. As the number of filtered values was 139, FSO line would fail to operate 139 hours per year. Resulting FSO line availability will be determined by Tab. 2.

One year has 8760 hours, or 8760 x 60 minutes.

For link availability of 99.99 %:

Unavailability is

$$1 - 0.9999 = 0.0001 (\text{outage}), \quad (1)$$

outage(%) = 1 - availability or

$$0.0001 \times 8760 \times 60 = 52.56 \text{ minutes} \quad (2)$$

Availability of FSO line is:

$$100\% - (139 \times 100\%) / 8760 = 100\% - 1.586\% = 98.41\% \quad (3)$$

TABLE 2
RELATIONSHIP BETWEEN SYSTEM AVAILABILITY AND
OUTAGE TIME

Availability %	Outage/Year
50	4380 hrs
80	1752 hrs
95	438 hrs
98	175 hrs
99	88 hrs
99.9	8.8 hrs
99.99	53 min
99.999	5.3 min
99.9999	32 sec

Availability of FSO line FlightStrata 155E in Košice is 98.41%. Availability of RF line is determined in literature [6]. Due to the fact that Slovakia is in rainfall area F, according to [6], availability of RF line is 99.9%. Consequential availability of hybrid FSO/RF is calculated according to statistical relations as follows:

Availability of hybrid FSO/RF system is defined as:

$$A_h = 1 - (1 - A_{FSO})(1 - A_{RF}) \quad (4)$$

where A_h is availability of hybrid line, A_{FSO} is availability of FSO line and A_{RF} is availability of RF line. Then :

$$A_c = 1 - (1 - 0.9841)(1 - 0.999) \quad (5)$$

$$A_c = 99.99841\% \quad (7)$$

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

IV. 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. Hybrid network, a combination of FSO main link and radio frequency back up link, is a viable solution to overcome the atmospheric attenuations.

Availability of FSO/RF link in Košice is 99.99841%. We are still collecting data of visibility for several years and availability can be determined more precise with more amount input data.

REFERENCES

- [1] F. Nadeem, V. Kvicera, A. M. Saleem, E. Leitgeb, S. S. Muhammad, G. Kandus, "Weather effects on Hybrid FSO/RF Communication link", *IEEE Journal on Selected Areas in Communications*, Vol. 27, No. 9, pp. 1687-1697, December 2009.
- [2] M. Alnaboulsi, H. Sizun, F. deFomel, "Fog Attenuation for Optical and Infrared Waves," *Journal of Optical Eng.*, Vol. 43, pp. 319-329, 2004.
- [3] R. Srinivasan, D. Sridharan, "The climate effects on line of sight (LOS) in FSO communication," *IEEE International Conference on Computational Intelligence and Computing Research*, 2010.
- [4] R. Luna, D. K. Borah, H. Tapse, "Behavior of hybrid optical/RF channels...", Proc. IEEE GLOBECOM 2009, paper# ONS-05/4, Honolulu, HA, 2009 Nov-Dec.
- [5] CH. Koh, "The Benefits of 60 GHz Unlicensed Wireless Communication," www.ydi.com
- [6] Wayne Pleasant Telaxis Communication, "Gigabit MM Wave Comm," *IEEE T802.16-02/02a*, 2002-03-12

Basic Nonlinear Model of Precision Gearbox

¹Matúš HRIC (3st year), ²Karol Kyslan (3nd year)
Supervisor: ³Viliam FEDÁK

^{1,2,3}Dept. of Electrical Drives and Mechatronics, FEI TU of Košice, Slovak Republic

¹matus.hric@tuke.sk, ²karol.kyslan@tuke.sk, ³viliam.fedak@tuke.sk

Abstract— The paper deals with basic introduction of operation principle the cycloid gearbox and torque transmission with the transformation component of gearbox. It deals with nonlinear fluence of angular transmission error and nonlinear hysteresis of high precision gearbox. In end of the paper is described a nonlinear mathematical model. All parasite nonlinearities are taking into consideration. The system analysis creates a starting point at design of position control system.

Keywords—Angular transmitting error, Cycloid gearbox, Gearbox hysteresis, Torsional stiffness

I. INTRODUCTION

Steadily increasing demands on precise control of position and movement of the processed material in various fields of industry put great emphasis to modern technologies, like to using more precise optical sensors, higher quality drive units, higher precision gears, faster and more powerful control processors reacting to all movement requirements. In the last time there are used compact servodrives (actuators) [1] which integrate the motor, power converter and gear into one corpus. Due to high requirements on precise positioning of machine equipment like robots, manipulators and machine tools, there were developed high precision gears (speed reducers) working on various principles. Understanding their principle of construction and their parasitic properties create one of basic parts at the servodrive analysis with the aim to compensate their adverse reactions with the goal to increase position control precision. The construction of low-speed servodrives with the possible lowest torque deviation is technically very intensive, a choice of a precision reducer with high gear ration and low parasitic values nowadays presents a standard at industrial robots manufacturing. This fact is extremely important at robots where any small angle deviation and nonlinear characteristic of the gear behavior increase with increasing length of the robot arm.

II. CYCLOIDAL GEARBOX

Cycloid gearbox presents one of two accurate gearboxes we analyze in the paper. In general, there are not many references about this kind of gearboxes. Description of its performance can be found in [2]. The principle of operation is similar to that of the planetary gearing but there is not any central wheel and instead of this there is a catcher system transmitting angular speed and torque by an eccentric. One of

possible construction solutions of the cycloid gearbox and torque transmission by the transformation component in a cross form is shown in Fig. 1. In shows a principal illustration of gear ration of a cycloid gearbox, where P is number of needles and L is number of teeth of a trochoidal cogging.

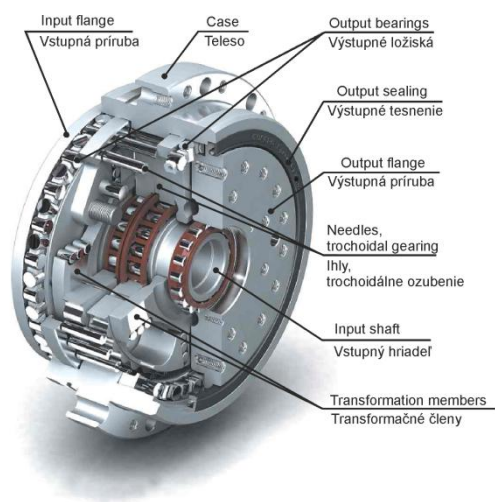


Fig. 1 Cycloid gearbox

For the gear ratio calculation the following formula can be derived (Fig. 2):

$$i = \frac{P - L}{L} \quad (1)$$

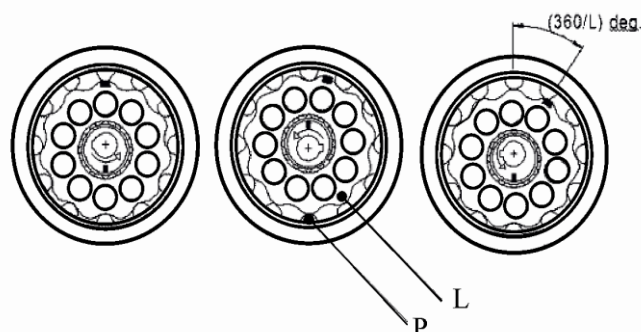


Fig. 2 Basic calculation of gear ratio of a cycloid gearbox

A disadvantage of a classic cycloid gearbox consists in a torque ripple at constant speed that physically can be explained using Fig. 3. Here rotation of two wheels is described as a mutual generating where axle of one wheel is displaced by an eccentricity [3]. In Fig. 3 r_{in} and r_c present radiuses of the geared wheels, e is the eccentricity value and

φ_{in} , φ_{out} are angles of geared wheels displacement.

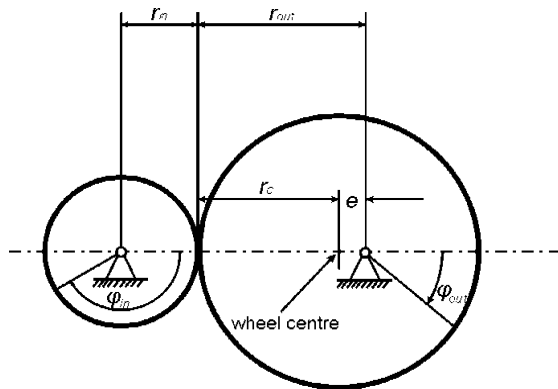


Fig. 3 Principle of transmitting torque ripple

Actually, relations in the gearbox are much more complicated. The gear cogged wheels are mutually displaced by 180° thanks to eccentricity shaft. Due to this bearing the torque ripple is linearized because both wheels are geared simultaneously displaced by 180°. Typical representatives of such gearboxes are these produced by company Spinea [3],[6] are characterized by smooth operation, high robustness, stiffness, and overloading.

Defects in wheel tooth or in annular gear bearing of the cycloid gearbox become evident as an angular transmission error. Based on this, the gear ration also depends on position of the cycloid wheel and the following formula is valid:

$$\dot{\varphi}_{in} = i(\varphi_{out}) \dot{\varphi}_{out} \quad (2)$$

$$M_a = \frac{M_{out}}{i(\varphi_{out})} \quad (3)$$

It is impossible to derive accurately an equation for the gear ratio $i(\varphi_{out})$ because manufacturing errors causing error of the angular transmission cannot be mathematically described by a simple equation. The angular transmission error is repeated in the next revolution and due its periodicity it can be written down by the Fourier series, as follows:

$$u \varphi_{out} = i + \sum_{j=1}^{n_u} A_j \cos(f_j \varphi_{out} + b_j) \quad (4)$$

where f is rotational frequency, i is the basic gear ratio with the amplitude A and b is a phase deviation angle of the gear ratio.

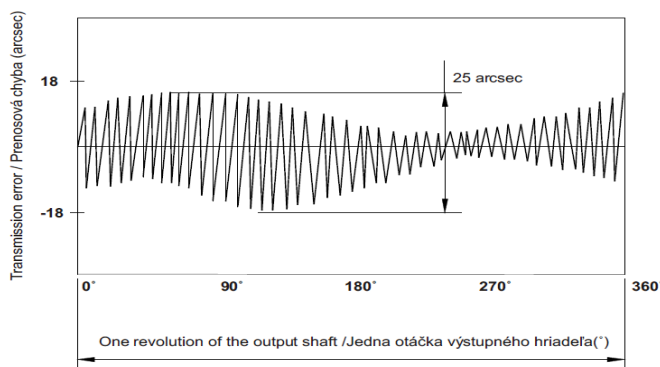


Fig. 4 Angular transmission error of the gearbox type SPINEA TS 140

Fig. 4 shows a measured angular transmission error depending on one revolution of the shaft.

III. HYTERESIS

Along with the angle transmission error, the hysteresis curve presents another characteristic feature of the gearbox. From the hysteresis curve the specialist can get important information about static and dynamic properties of the gears. Gear stiffness, tooth backlash, and information about position accuracy belong to the most meaningful variables that can be obtained from measured hysteresis characteristics. In general, the hysteresis is an undesirable factor because it influences gear accuracy and wear. The gear hysteresis and by this influenced the gear load characteristics can be measured as follows: The hysteresis curve is obtained from the measurement where the input shaft and body of the gear are fixed against the rotation and the torque actuates through input (driving) flange in both directions. [4]

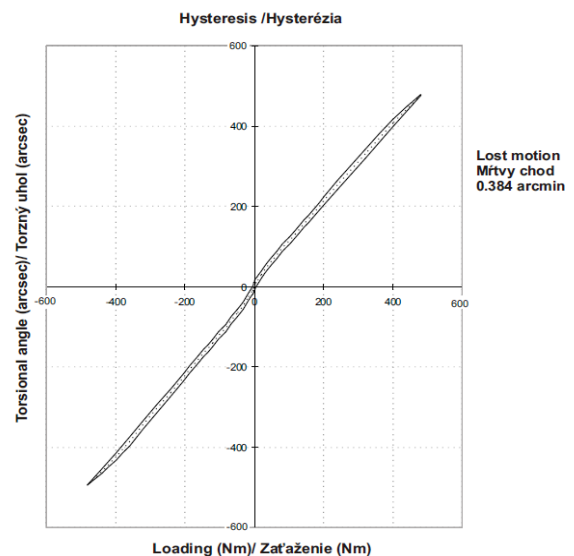


Fig. 5 Hysteresis curve TS 140-139-TB

In the ref. [5] there is described a mutual relationship between hysteresis curves of the gear from which effectiveness of gear torsional stiffness η_t can be calculated as follows:

$$\eta_t = 1 - \frac{E_h}{2E_c} \quad (5)$$

where E_h is energy of losses and a E_c is so called Feder energy for which the following equations are valid:

$$E_H = \left| \int_{-\Delta\varphi_{max}}^{\Delta\varphi_{max}} M_1 \Delta\varphi d\varphi - \int_{-\Delta\varphi_{max}}^{\Delta\varphi_{max}} M_2 \Delta\varphi d\varphi \right| \quad (6)$$

and

$$E_c = \int_{-\Delta\varphi_{max}}^{\Delta\varphi_{max}} M_c \Delta\varphi d\varphi \quad (7)$$

where M_1 , M_2 are torque values of the upper and lower branch of the hysteresis curve and M_c corresponds to a lossless elastic torque.

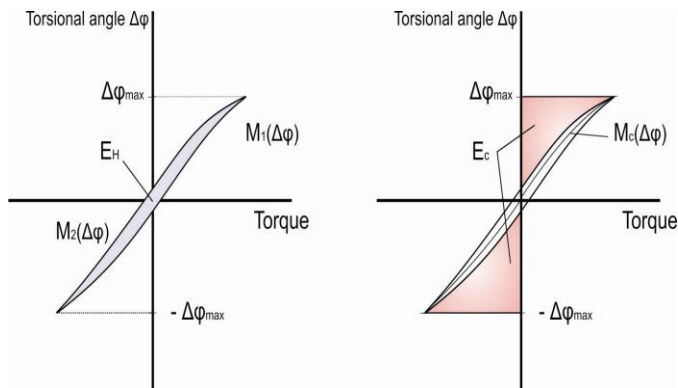


Fig. 6 Hysteresis curve characteristic variables

The method of effectiveness of gear torsional stiffness cannot be considered for effectiveness of the gear as a whole. Effectiveness of gear torsional stiffness presents an important measure for determination of gear internal losses that cause increased wearing or leads to excessive thermal stress of single components.

IV. MODEL AND SIMULATION RESULTS

A general model of a precise gear, valid for large class of gears regardless their construction shown in Figure 8. It can be described as two-mass system in Figure. 7 with backlash, elastic connection, and friction and without considering angle transfer error. The torsional stiffness is enough high what is a characteristic parameter for precise gears.

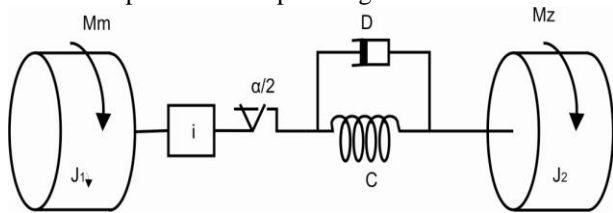


Fig. 7 Kinematic scheme of two-mass gear system with backlash

Fig. 8 shows the block diagram derived from the mathematical model.

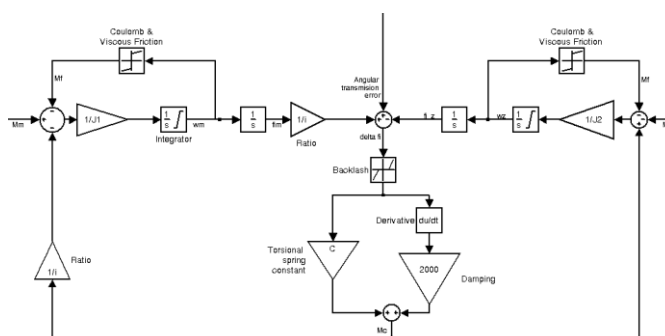


Fig. 8 Block diagram of the gear with backlash, friction and angular transmitting error

For simulation we used the following parameters (valid for the gear of Spinea type TS70): $i = 57$, $C = 101000$ Nm/rad, $D = 2000$ Nms/rad, backlash (lost motion) $\alpha = 0,0004363$ rad $\approx 1,5$ arcmin [3]. Further parameters: $J_1 = J_m + J_1 = 0,03$ kgm², $J_2 = 0,102$ kgm², Coulomb friction coefficient = 0,003, viscous

friction coefficient = 0,001. In the following figures are the time and position responses of mathematical model of cycloid gearbox. Fig.9 show the position nonlinearity response of the angular transmission error of cycloid gearbox. Fig.10 show the position response $\Delta\phi$ dependence ϕ_{out} in gearbox model,

Fig. 11 show time response of desired position output shaft and position output shaft with angular transmission error of gearbox model. Fig 12 show the detail of Fig 11.

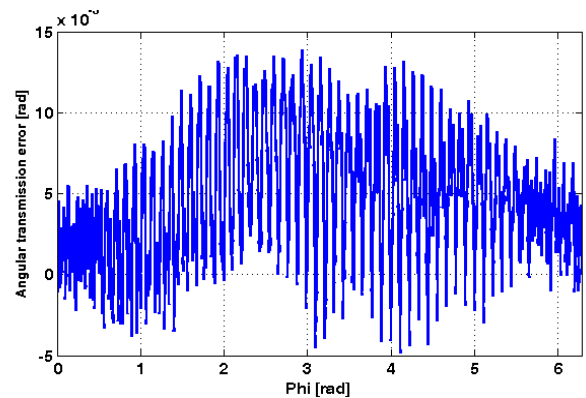


Fig. 9 Position response of angular transmission error of cycloid gearbox

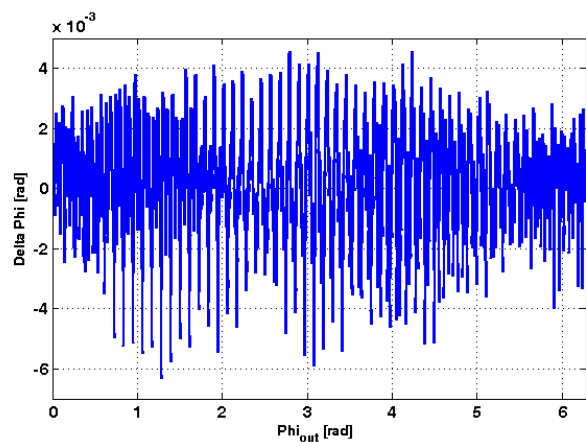
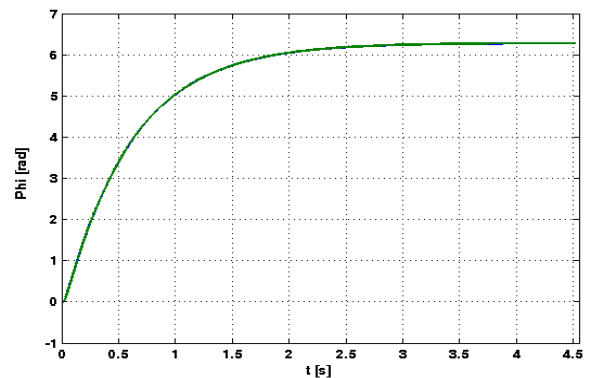

 Fig. 10 Position response $\Delta\phi$ in dependence ϕ_{out} in gearbox model


Fig. 11 Time response of desired position output shaft and position output shaft with angular transmitting error in gearbox model

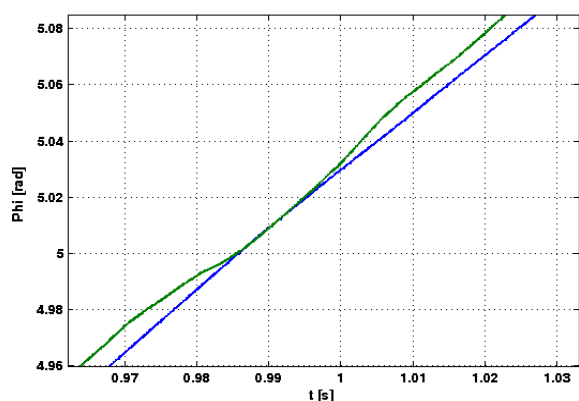


Fig. 12 Detail of Figure 11.

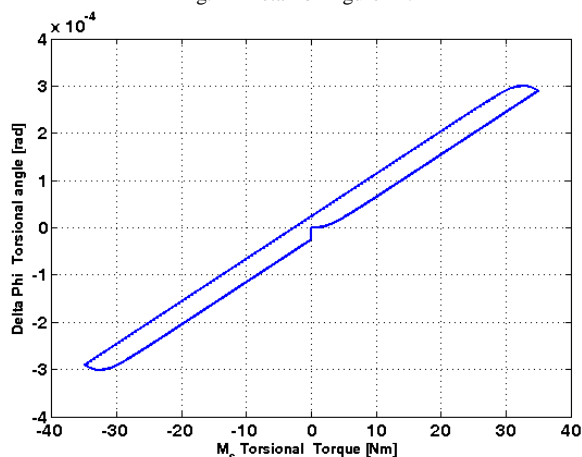


Fig. 13 Hysteresis curve of gearbox model

V. CONCLUSION

Understanding of operation and analysis nonlinear properties of a gear presents starting point at designing control structures of the high precision servodrives used in robotics. Analysis of specific nonlinearities of the actuator with the following compensation creates essential part of position control. The general model of precision gearbox and the responses results is needed for verification of real system. This system will be measured in the future.

REFERENCES

- [1] R. Jánoš, J. Semjon "Trendy uplatnenia MLR a A v jednotlivých technológiách, výrobných oboroch. In: Transfer inovácií. Č. 18 (2010), s. 191-194. - ISSN1337-7094.
- [2] M. Lehmann, "Berechnung und Messung der Kräfte in einem Zykloiden-Kurvenscheiben-Getriebe", Technische Universität München, diss., 1976..
- [3] M. Thümmel "Modellbasierte Regelung mit nichtlinearen inversen Systemen und Beobachtern von Robotern mit elastischen Gelenken", Technische Universität München, Dissertation 2006.
- [4] Catalogue SPINEA, Ltd, TwinSpin, vol. I/2009, pp. 58, 2009 ch. 4.
- [5] T. Kalender, "Statistische Modellierung von Präzisionsgetrieben in elektromechanischen Antriebssystemen." Fortschritt-Berichte, VDI, Reihe 1, Nr. 236, 1994, VDI-Verlag, ISBN 3-18-323601-X, 236 S.
- [6] M. Balara, "Dynamické charakteristiky ložiskového reduktora", In Automatizácia a riadenie v teórii a praxi, ARTEP 2008, Workshop odborníkov z univerzít, Vysokých škôl a praxe 20. 2 – 22. 2 2008, Stará Lesná, SR

ACKNOWLEDGMENT

This publication is the result of the project implementation: Výskum modulov pre inteligentné robotické systémy, ITMS: 26220220141, supported by the Research & Development Operational Programme funded by the ERDF.

This work was also supported by the Slovak Research and Development Agency under the contract No. APVV-0185-10.



Cambridge Correlator in Traffic Sign Recognition System

¹Tomáš HARASTHY (2nd year), ²Pavol MIŠENČÍK (3rd year)
Supervisor ³Ján TURÁN

^{1,2}Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

¹tomas.harasthy@tuke.sk, ²pavol.misencik@tuke.sk, ³jan.turan@tuke.sk

Abstract— This paper present proposal, experiments and results of Traffic Sign Recognition System based on Optical Correlator. In this system Cambridge Correlator was used. Traffic Sign Recognition System consists of three main blocks, Preprocessing, Optical Correlator and Traffic Sign Identification. Real traffic scene is captured by a color camera. Key frame is extracted and then is preprocessed in preprocessing block. The Region of Interest (ROI) is defined. Preprocessed ROI goes to Optical Correlator and then is compared with reference images (traffic signs). Position of Correlations peaks in correlation plane presents position of traffic sign in input plane. Several experiments were done with proposed system. Results and conclusions are discussed.

Keywords—Color segmentation, Correlations peaks, Optical Correlator, Traffic Sign Identification,

I. INTRODUCTION

There are many types of video based object recognition systems to recognize Traffic Signs as a part of complex Drive Assistance System (DAS) [1-3]. Each of these partial systems used in DAS has some specific requirements but one of the most important is real time recognizing. Video processing considers very high speed of processing. There are many Traffic Sign Recognition Systems mostly based in Shape detection and Color segmentation such as Automatic Detection and Classification of Traffic Signs, Real Time Road Signs [3] and Traffic Sign Recognition System [3]. Traffic Sign recognition is very important part of DAS [1-3]. It can save drivers money (if driver overlook some traffic sign), but also, what is more important, can save a drivers and crews life. DAS generally consists of more than one system. There are Park Assistance System, Active Park Assistance, Low Speed Safety and many other systems. In this paper is presented one of these subsystems, The Traffic Sign Recognition System.

The proposed Traffic Sign Recognition System is based on Optical Correlator, specifically Cambridge Correlator. This type of optical correlator use Optical Fourier Transform to make the Joint Power Spectrum and then optical correlation between two input images. Cambridge correlator belongs to family of correlators called Join Transform Correlators, (JTC). JTC produce whole optical correlation in real time so all time restriction are in preprocessing of input image and are looking for Traffic Sign in input image (video).

In Chapter 2 the main information about used Cambridge

Correlator is presented.

In Chapter 3 the proposed system will be presented. There is shown hardware diagram of used hardware components. In Chapter 3 is also block diagram of whole proposed system.

In Chapter 4 the experiments and results will be presented. The experiments were done with static images captured in real traffic in Slovakia. Experiments have been done with chosen traffic signs and results of recognition are recorded in tables.

Conclusions are covered in Chapter 4.

II. CAMBRIDGE CORRELATOR

Cambridge Correlator is device to compare two images or two-dimensional data in real time. Image data that enter to the optical system are compared during correlation process in terms of similarity and relative position. Comparing is done between a reference image and input image. Reference image could be some image from database and input image is image captured by external camera [6].

Optical system of Cambridge Correlator is based on Fourier Transform Engine© - a compact, yet hugely powerful processing system [5,6], built on the diffraction principles of the Optical Fourier Transform (OFT). Its new “W” design allows the electro-optical components to operate at their full potential. The optical output consists of highly localized intensities; know as correlation peaks, relating to wherever a strong match has been found between the compared images [5]. The intensity I , of the peaks provides a measure of the similarity of the images being compared, whilst the position (x,y) , of the peaks in the output denotes how the images are relatively aligned in the input scene. The main parts of Optical Correlator are shown on Fig. 1.

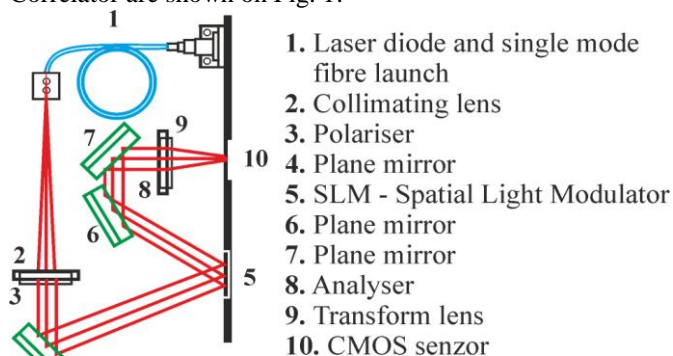


Fig. 1. Basic components of Optical Correlator [5].

III. TRAFFIC SIGN RECOGNITION SYSTEM

Nowadays Traffic Sign Recognition System is very popular topic in most of designers of cars. This system helps keep car and crew safety. Process time for this system is very important so using Optical Correlator in these systems can help to obtain much better results than system based only on software applications. Cambridge Correlator described above is used in our system. This system uses Optical Fourier Transform to obtain correlation between input data. This correlation is done in real time, so process time mostly depends on Preprocessing and Identification of Traffic Sign. [2]

A. Hardware Scheme of the System

In Fig. 2 is shown hardware scheme of designed Traffic Sign Recognition System consists of: Color Camera, Computer Unit, Optical Correlator and Display.

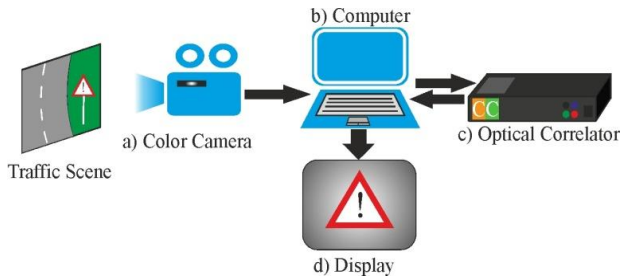


Fig. 2. Hardware scheme.

Color camera captures the road all time and is connected to the computer unit. Using preprocessing software, we can remove irrelevant samples from input scene and keep just relevant information for traffic sign recognition. For example, we can keep just red bit-map from input scene so there will be just objects with red color (include traffic signs). After preprocessing the preprocessed traffic scene enters to the optical correlator. Then the correlation between preprocessed input and database of traffic signs occurs. Information about correlation enters back to the computer unit and result of comparing is shown on display. On display we can see the traffic sign which was recognized.

B. Block Scheme of Traffic Sign Recognition System

In Fig. 3 is shown block diagram of proposed system to recognize Traffic Signs using Cambridge Correlator. System consists of three main blocks: Preprocessing, Optical Correlator and Traffic Signs Interpretation.

Traffic scene is captured with the CMOS color camera. From traffic video stream the key frame is extracted and then used in next processing. First step of preprocessing is correction of exposure. In next step of preprocessing, the color filter is used. Color filters are designed to remove all irrelevant colors in input traffic scene. Input image is in HSV color space. The color segmentation is processed. Every traffic sign has his dominant color. On the Slovak roads most often Yellow, Red and Blue color are used. This means we need to create three binary maps, one for each of these colors [4]. By analyzing hue component (H), we can identify Blue, Yellow and Red regions in our processed image. For each image pixel is possible to use hue-based detection for Blue, Red and Yellow color. For each color following equation are used

$$Y = e^{\frac{-(x-42)^2}{30^2}} \quad (1)$$

$$R = e^{\frac{-x^2}{20^2}} + e^{\frac{-(x-255)^2}{20^2}} \quad (2)$$

$$B = e^{\frac{-(x-170)^2}{30^2}} \quad (3)$$

Equations Y gives values close to 1 for Yellow regions, R gives values close to 1 for Red regions and B gives values close to 1 for Blue regions. In this equations we can see, that H can be from range 0-255. Yellow can be detected near value 42, Red near values 0 and 255 and Blue value is 170. These equations can be tuned for every other needed color. After threshold for detected colors three binary maps (Red, Blue and Yellow) are created. Color filtering removes all irrelevant colors. Color filtering get image with color clusters, which are irrelevant for detecting Traffic Sign. In next step in preprocessing irrelevant color clusters are removed. To small regions and too big regions are discarded. In first step white point in image is found [4]. Searching is done by rows.

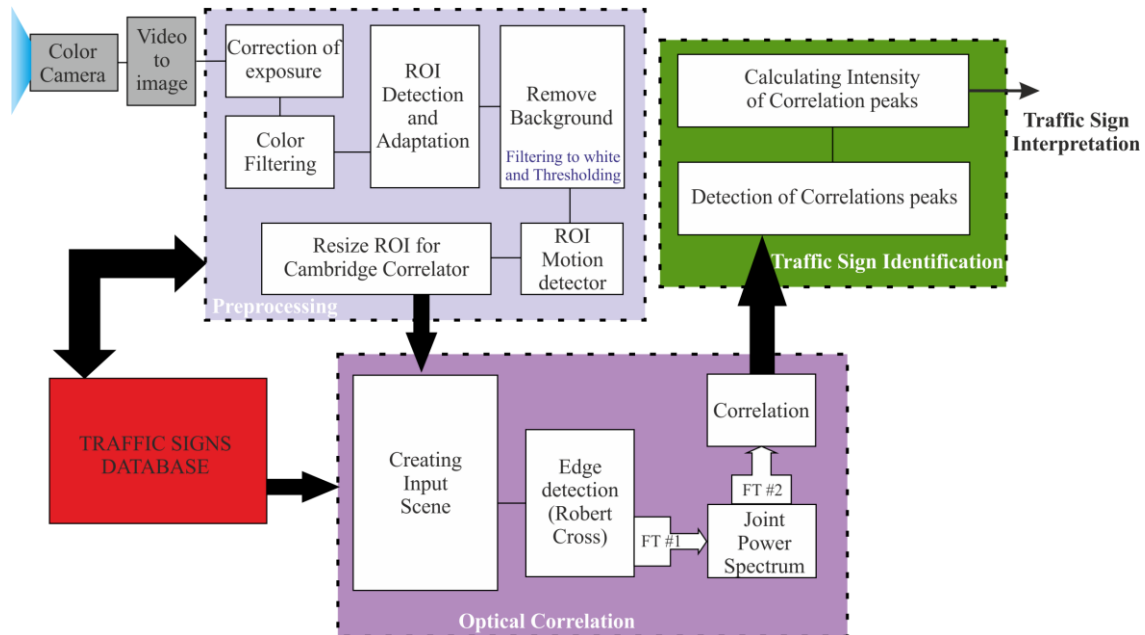


Fig. 3. Block diagram of Traffic Sign Recognition System.

After finding first white point, method called seed-fill is used. With this method we are finding regions. After color filtering and irrelevant cluster removing, the potential Region of Interest (ROI) is extracted. The potential ROI is defined at the input image. The ROI Motion Detector using moving vectors to discard moving features from traffic sign on moving object, i.e. vehicle.

IV. EXPERIMENTS AND RESULTS

In our experiment with designed traffic sign recognition system, are four types of traffic signs, “STOP”, “CROSSWALK”, “MAIN ROAD” and “YIELD”, showed in Fig. 4 compared. For this study traffic scenes from road of Slovakia in daytime lighting conditions are taken, in urban traffic environment.



Fig. 4. Recognized Traffic Signs.

If the background of the processed image is deleted, ROI image containing possible traffic sign is obtained.

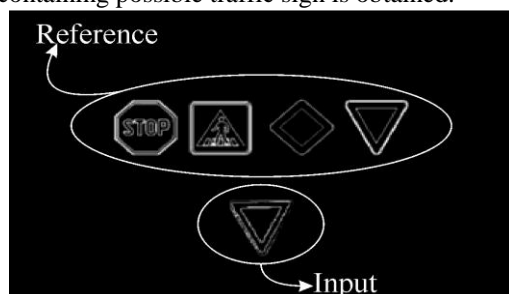


Fig. 5. Input plane in correlation process.

To minimize energy of DC noise in Optical Correlators Joint Power Spectrum (JPS) we can use some edge detection in ROI image. In our experiments simple Roberts Operator is used [4]. In Fig. 5 we can see input ROI image preprocessed with Roberts Operator. In top row are aligned reference signs and in a bottom is recognized sign.

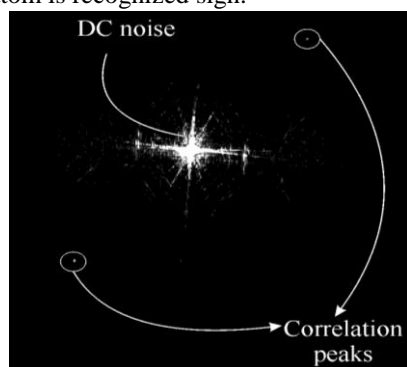


Fig. 6. Correlation.

Intensity of obtained peaks is represented in 8 bit depth grayscale, so every peak can take values from 0 to 255. Correlation peaks has exactly defined area of occurrence in correlation plane. Thanks these areas we can determine which traffic sign was recognized. JPS is Optical Fourier Transformed to obtain correlation peaks, showed in Fig.6. In the middle of figure is situated DC noise. Correlations peaks are situated according the recognized signs are aligned

in the input scene. In the Tab.1 are results of recognizing tested Traffic Signs. There is information about intensities of each correlations peak.

TABLE I
RESULTS OF TESTED RECOGNITION

Traffic Sign	INTENSITY OF 1. PEAK	INTENSITY OF 2. PEAK	AVERAGE INTENSITY [%]
YIELD 1	I1 = 116	I2 = 109	44
YIELD 2	I1 = 70	I2 = 41	22
YIELD 3	I1 = 158	I2 = 158	62
YIELD 4	I1 = 65	I2 = 62	25
Main Road 1	I1 = 124	I2 = 83	41
Main Road 2	I1 = 247	I2 = 239	95
Main Road 3	I1 = 53	I2 = 94	29
Main Road 4	I1 = 97	I2 = 56	30
Crosswalk 1	I1 = 208	I2 = 236	87
Crosswalk 2	I1 = 233	I2 = 212	87
Crosswalk 3	I1 = 143	I2 = 162	60
Crosswalk 4	I1 = 218	I2 = 193	80
STOP 1	I1 = 191	I2 = 176	72
STOP 2	I1 = 135	I2 = 195	65
STOP 3	I1 = 157	I2 = 170	64
STOP 4	I1 = 161	I2 = 249	80

From Table 1 we can see that the highest intensities values were obtained with CROSSWALK and STOP. These traffic signs contain more details, than YIELD and MAIN ROAD. Thanks these details the higher intensity of correlation peak was detected.

V. CONCLUSION

In this paper the Traffic Signs Recognition System with using Optical Correlator was presented. Traffic scenes were captured with color CMOS 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. Experimental results demonstrate the main advantage of this system is that using optical correlator; it operates at very high speed compared to other pure software implementations running on PC. In future we want to focus on preprocessing methods of input information and use preprocessed video as input for Optical Correlator. Our focus will be also on implementation of some preprocessing algorithms in hardware using FPGA.

ACKNOWLEDGMENT

This work is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] M. Ahmadi and H. Adeli, “Enhanced Probabilistic Neural Network with Local Decision Circles” in *A Robust Classifier, Integrated Computer-Aided Engineering*, Vol. 17, no. 3, 2010, 197-210.
- [2] T. Harasthy, J. Turan, L. Ovsenik, K. Fazekas, “Traffic Signs Recognition with Using Optical Correlator”, in *IWSSIP 2011*, Sarajevo, Bosnia and Herzegovina, 16-18 June 2011, 239-242. ISBN: 978-9958-9966-1-0, IEEE Catalog number: CFP1155E-PRT.
- [3] M. Fifik, J. Turán, L. Ovsenik and K. Fazekas, “Experiments with a Transform based Traffic Sign Recognition System”, in *Of 17th IWSSIP 2010*, Rio de Janeiro, Brazil, June 17-19, 2010, 227-230.
- [4] S. E. Umbach, *Computer Vision and Image Processing and A Practical Approach Using CVIptools*, Prentice Hall PTR, Upper Saddle River, NJ 07458, June 1999.
- [5] Cambridge Correlators, <http://www.cambridgecorrelators.com>, 2012.
- [6] P. Mišenič, „Optický signálový procesor“, *Diplomová práca*, 2009

Comparison of Motor Dynamics of DC and BLDC Motors

¹Tibor BALOGH
Supervisor: ²Viliam FEDÁK

^{1,2}Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

¹tiborb@gmail.com, ²viliam.fedak@tuke.sk

Abstract—Paper describes comparison of two speed controlled DC motor drives from view of their dynamics and robustness against the rotor temperature changes: a field-oriented brushless DC motor drive and brushed DC motor. The control models in the MATLAB/Simulink environment were developed by applying mathematical equations both motors and using common design control methods. Analysis of time responses of both controlled drives shows a better performance of the drive system with the BLDC motor that the drive with the DC motor of a comparable size.

Keywords— brushless DC motor, control algorithm, DC motor, simulation

I. INTRODUCTION

Brushed DC (BDC) motors with commutator are widely used in applications ranging from toys to these ones used to drive industrial lines. The BDC motors are inexpensive, easy to drive, and are readily available in all sizes and shapes. [1] The Brushless Direct Current (BLDC) motor is rapidly gaining popularity by its utilization in various industries, such as Appliances, Automotive, Aerospace, Consumer, Medical, Industrial Automation Equipment and Instrumentation. As the name implies, the BLDC motors do not use brushes for commutation; instead, they are electronically commutated. The BLDC motors have many advantages over BDC motors and induction motors (IM) [1] that should be mentioned: better speed versus torque characteristics, high dynamic response, high efficiency, long operating life, noiseless operation, higher speed ranges

In addition, the ratio of torque delivered to the size of the motor is higher, making it useful in applications where space and weight are critical factors, [2].

II. CONSTRUCTION AND OPERATING PRINCIPLE

All BDC motors consist of the same basic components: a stator, rotor, brushes and a commutator. The stator generates a stationary magnetic field that surrounds the rotor. This field is generated by either permanent magnets or electromagnetic windings. The different types of BDC motor are distinguished by construction of the stator or the way the electromagnetic windings are connected to the power source. The rotor, also called the armature, is made up of one or more windings. When the windings are energized they produce a magnetic field. The magnetic poles of this rotor field will be attracted to

the opposite poles generated by the stator, causing the rotor to turn. As the motor turns, the windings are constantly being energized in a different sequence so that the magnetic poles generated by the rotor do not overrun the poles generated in the stator. This switching of the field in the rotor windings is called commutation.

The BLDC motor is also referred to as an electronically commutated 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, [2]

The simple brush DC and brushless DC motor construction is shown in Fig. 1.

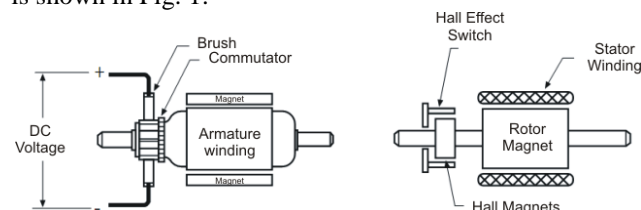


Fig. 1. BDC and BLDC motor construction

In the BDC motor, the current polarity is reversed by the commutator and the brushes, but in the BLDC motor, the polarity reversal is performed by semiconductor switches which are to be switched in synchronization with the rotor position. Besides of the higher reliability, the missing commutator brings another advantage. The commutator is also a limiting factor in the maximal speed of the BDC motor. Therefore, the BLDC motor can be employed in applications requiring high speed, [1], [3].

Replacement of a BDC motor by a BLDC motor place higher demands on control algorithm and control circuit. Firstly, the BLDC motor is usually considered as a three-phase system. Thus, it has to be powered by a three-phase power supply. Next, the rotor position must be known

at certain angles, in order to align the applied voltage with the back-EMF. The alignment between the back-EMF and commutation events is very important. In this condition the BLDC motor behaves as a BDC motor and runs at the best working point. But the drawbacks of the BLDC motor caused by necessity of power converter and rotor position measurement are balanced by excellent performance and reliability, and also by the ever-falling prices of power components and control circuits. The simple motor model of a BLDC motor drive consisting of a 3-phase power stage and a BLDC motor is shown in Fig. 2.

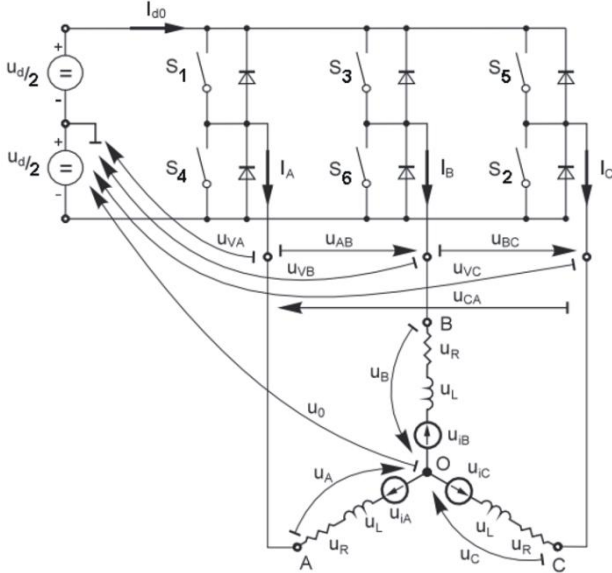


Fig. 2. BLDC motor circuit model

III. MATHEMATICAL MODELS OF THE BDC AND BLDC MOTORS

A. Model of the BDC Motor

The mathematical model of the BDC motor is expressed as follows. The basic equations of the motor are generally known:

– the rotor circuit:

$$v = Ri + L \frac{di}{dt} + e \quad (1)$$

– the back-EMF:

$$E = K_e \omega \quad (2)$$

– the motor torque:

$$T_m = K_m i \quad (3)$$

– the mechanical equation:

$$T_m - T_l = J \frac{d\omega}{dt} + b\omega \quad (4)$$

where V is terminal voltage [V], R – armature resistance [Ω], L – armature inductance [H], e – motor back-EMF [V], i – armature current [A], K_e – back-EMF constant [V/rad.s⁻¹], K_m – the motor torque constant [Nm/A], T_l – load torque [Nm], J – rotor inertia [kgm²], b – friction constant [Nms.rad⁻¹].

B. Model of the BLDC motor

The model of each phase of BLDC motor armature winding is expressed similarly like at BDC motor:

$$v_a = Ri_a + L \frac{di_a}{dt} + e_a \quad (5)$$

$$v_b = Ri_b + L \frac{di_b}{dt} + e_b \quad (6)$$

$$v_c = Ri_c + L \frac{di_c}{dt} + e_c \quad (7)$$

where L is armature self-inductance [H], R – armature resistance [Ω], v_a, v_b, v_c – terminal phase voltage [V], i_a, i_b, i_c – motor input current [A], and e_a, e_b, e_c – motor back-EMF [V].

In the 3-phase BLDC motor, the back-EMF is related to a function of rotor position. The back-EMF of each phase has 120° phase angle difference so equation of each phase is as follows:

$$e_a = K_w f(\theta_e) \omega \quad (8)$$

$$e_b = K_w f(\theta_e - 2\pi/3) \omega \quad (9)$$

$$e_c = K_w f(\theta_e + 2\pi/3) \omega \quad (10)$$

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:

$$\theta_e = \frac{p}{2} \theta_m \quad (11)$$

where θ_m is mechanical rotor angle [rad].

Total motor torque is represented as summation of that of each phase:

$$T_e = \frac{e_a i_a + e_b i_b + e_c i_c}{\omega} \quad (12)$$

where T_e is total torque output [Nm], p – number of pole pairs.

The equation of mechanical part is represents as follows:

$$T_e - T_l = J \frac{d\omega}{dt} + b\omega \quad (13)$$

where T_l is load torque [Nm], J – rotor inertia [kgm²], b – friction constant [Nms.rad⁻¹].

IV. FIELD-ORIENTED CONTROL OF THE BLDC MOTOR

The traditional control method for BLDC motors drives the stator in a six-step process, which generates oscillations on the produced torque. In the six-step control, a pair of windings is energized until the rotor reaches the next position, then the motor is commutated to the next step. Hall sensors determine the rotor position electronically commutate the motor. Advanced sensorless algorithms use the back-EMF generated in the stator winding to determine the rotor position.

The dynamic response of six-step control (also called trapezoidal control) is not suitable e.g. for washing machines because the load is changing dynamically within a wash cycle. Further, in a front load washing machine, the gravitational power works against the motor load when the load is on the top side of the drum. Only advanced algorithms such as field-oriented control (FOC) can handle these dynamic load changes where the BLDC motor behaves as a BDC motor.

FOC is a math-intensive technique for controlling BLDC

and AC induction motors. Power-factor correction (PFC) and FOC save energy and quiet rotating machinery, [2].

By the vector control of the BLDC motor, we can satisfy the condition in which $i_d = 0$ and $i_q = I_m$, at preserving the angle of $\Theta = 90^\circ$. In the fact, at the vector control of the BLDC motor we control the stator current torque component. The flux linkage component of the stator current is zero. Fig. 3 shows the above-mentioned condition.

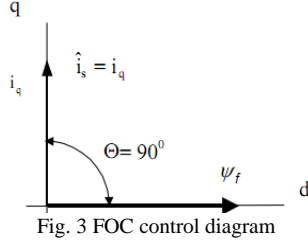


Fig. 3 FOC control diagram

Transformations from stator coordinate system (zero rotating speed) to the synchronously rotating coordinate system are significant for vector control. Coordinates of two-phase systems in the complex plane plot are shown in the Fig. 4 where $\{\alpha, \beta\}$ is the coordinate system connected with the stator, $\{d, q\}$ – with the rotor and $\{x, y\}$ – with the magnetic field (vector of rotating magnetic flux).

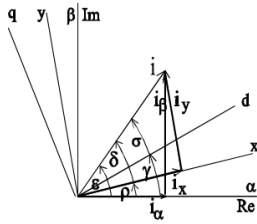


Fig. 4 Complex plane of coordinate systems [3]

After the PI controllers we have two voltage vectors in the rotating $d-q$ axes. To get back to the 3-phase motor voltage we need to perform the complementary inverse transform. Firstly we transform the variables from two-axis rotating $d-q$ frame into the two-axis stationary frame $\alpha-\beta$. The voltages V_d and V_q are transformed back into the stationary reference frame using the transformation angle to obtain quadrature voltage values V_α and V_β (Inverse Park Transform):

$$(10)$$

The next step consists in transformation from the stationary two-axis $\alpha-\beta$ frame into the stationary three-axes, 3-phase reference frame of the stator. V_α and V_β are mathematically transformed back into three phase-voltages V_a, V_b, V_c , which determine the new PWM duty-cycle in the converter. Mathematically, this transformation is accomplished with Inverse Clark Transform, as matrix is shown below:

$$\begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} = \frac{2}{3} \begin{bmatrix} 1 & -1/2 & -1/2 \\ 0 & \sqrt{3}/2 & -\sqrt{3}/2 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} V_\alpha \\ V_\beta \\ 0 \end{bmatrix} \quad (11)$$

Fig. 5 shows the control scheme of the field-oriented control of the BLDC motor. The model was developed from the above-mentioned coordinate transformations.

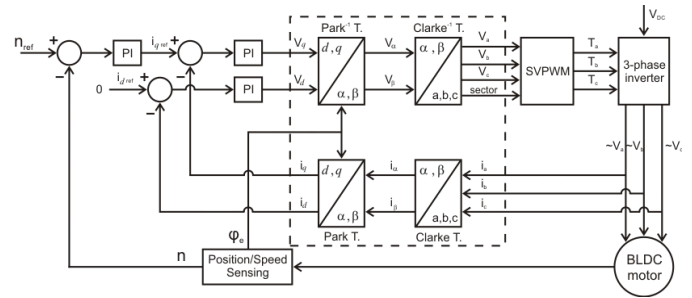


Fig. 5 Control scheme of the FOC control method

V. SIMULINK MODEL OF THE DC AND BLDC MOTORS

Unlike a BDC motor, the commutation of a BLDC motor is controlled electronically. It is important to know the rotor position in order to understand which winding should be energized following the energizing sequence. However, one electrical cycle may not correspond to a complete mechanical revolution of the rotor. The number of electrical cycles to be repeated to complete a mechanical rotation is determined by the rotor pole pairs. The number of electrical cycles/rotations equals the rotor pole pairs [3], [4]. The commutation sequences are listed in Table I.

TABLE I.
ELECTRICAL DEGREE, HALL SENSOR VALUE AND CORRESPONDING COMMUTED PHASE IN CLOCKWISE ROTATION OF THE ROTOR.

Electrical degree	Hall sensor value (ABC)	Phase	Switches
0° - 60°	101	A-C	S1-S2
60° - 120°	001	B-C	S2-S3
120° - 180°	011	B-A	S3-S4
180° - 240°	010	C-A	S4-S5
240° - 300°	110	C-B	S5-S6
300° - 360°	100	A-B	S6-S1

Fig. 6 shows the simulation model of the vector control of the BLDC motor that was developed by transformation in the coordinate systems and mathematical equations which are mentioned above (the detailed MATLAB/Simulink model of the three-phase converter is generally known and due to this fact it not mentioned here).

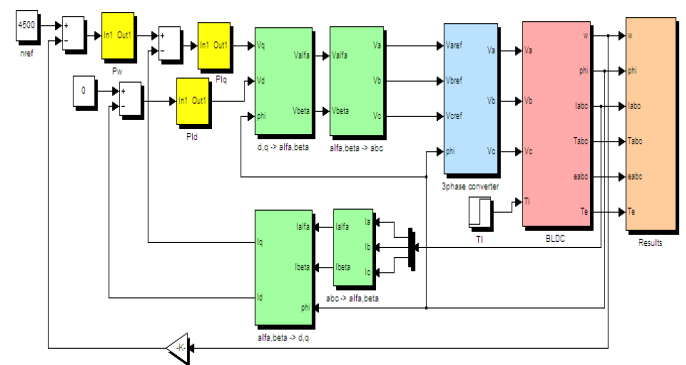


Fig. 6 Simulation model of the control system of BLDC motor

Fig. 7 shows the classical model of the control of the BDC motor with cascade speed and current controllers to be compared with previous one.

VI. CALCULATION OF THE BDC DRIVE CONTROLLERS

A controller calculates an error value as the difference between a measured and a desired setpoint. The controller attempts to minimize the error by adjusting the process control inputs.

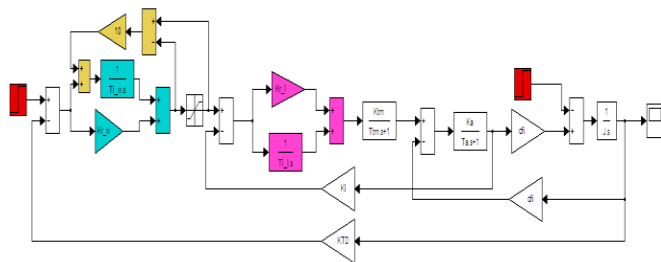


Fig. 7 Simulation model of the control system of BDC motor

The calculation algorithm for the current controller is based on the optimal module criterion. The gain K_r and time constant T_r is calculated by:

$$(12)$$

$$(13)$$

The calculation algorithm for the speed controller is based on the symmetrical optimum criterion. The gain constant is the same like for the current controller. The time constant is calculated by:

$$(14)$$

where T_1 , T_2 are time constant of the controlled system, K_I is gain of the motor armature.

VII. SIMULATION RESULTS

The coordinate transformations and suitable controller parameters play a very important role in the motion control. With incorrect transformations and controller parameters it is not possible to control the BLDC motor. The motor control is highly influenced by the transients. With increasing a temperature of the motor is changing the resistance of the armature windings of the motor. Change of the resistance has an effect on transients. Fig 7 is shown an effect of the resistance change to the transients for the BLDC motor.

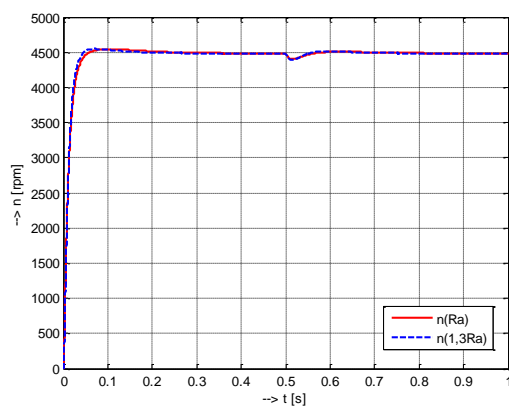


Fig. 7 Time response of the BLDC motor speed

Fig. 7 shows results of simulation of a BLDC motor with the following parameters: $V_d = 50$ V, $R = 4,98$ Ω , $L = 5,05$ mH, $p = 2$, $J = 18,17 \cdot 10^{-6}$ kgm², $B = 0$ Nms/rad, $k_w = 56,23 \cdot 10^{-3}$ V/rad.s⁻¹, $T_l = 0.2$ Nm. The load time is 0,5 s.

In Fig. 8 there is shown an effect of the resistance change to the transients for the BDC motor with the following parameters: $V_d = 420$ V, $R = 0,139$ Ω , $L = 3,56$ mH, $J = 2,2$ kgm², $B = 0$ Nms/rad, $T_l = 1230$ Nm. The load time is 0,5 s.

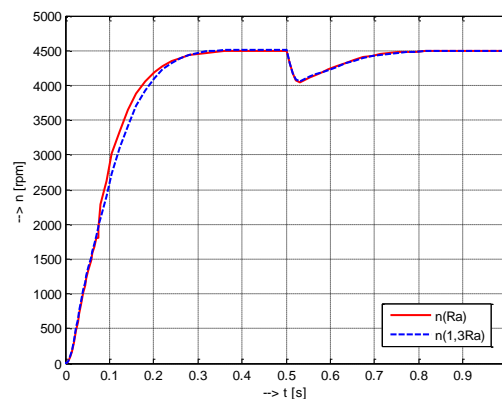


Fig. 8 Time response of the BDC motor speed

VIII. CONCLUSION

Simulation models of the DC drives with BDC and BLDC motors are presented here and their performance from view of system dynamics is compared. The simulation results show that the modelling is very useful in studying the drive system. Comparing the time course in Fig. 7 and Fig. 8 we can observe the difference between two motor dynamics. Dynamic and speed start of the BLDC motor is higher as by the BDC motor. The speed of the BLDC motor was decreased by the torque load only a little because of the better dynamic response. Change of the armature resistance has an effect on dynamics and transients of the mentioned BDC and BLDC motors.

ACKNOWLEDGMENT

The work was supported by Slovak Cultural and Educational Agency of the Ministry of Education of Slovak Republic under the contract KEGA 042TUK-4/2012 "Teaching Innovation in Control of Mechatronic Systems"

REFERENCES

- [1] P. Yedamale, "Brushless DC (BLDC) Motor Fundamentals", Microchip Technology Inc., 2003.
- [2] P. 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] B. Indu Rani, A. Mary Tom, "Dynamic Simulation of Brushless DC Drive Considering Phase Commutation and Backemf Waveform for Electromechanical Actuator", IEEE TENCON 2008, Hyderabad. ISBN: 978-1-4244-2408-5.
- [5] 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.
- [6] 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.
- [7] R. Krishnan, "Permanent Magnet Synchronous and Brushless DC Motor Drives". CRC Press, Taylor & Francis Group, New York 2009.
- [8] A. Consoli, S. Musumeci, A. Raciti and A. Testa, "Sensorless Vector and Speed Control of Brushless Motor Drives". IEEE Trans. on Industrial Electronics, vol. 41, pp. 91-96, February 1994.
- [9] J. Michalik, J. Buday, "Electrical Machines", EDIS Publishers in Zilina, 2006. ISBN: 80-8070-568-2, in Slovak.
- [10] L. Zboray, F. Durovsky, J. Tomko, "Controlled Drives". VIENALA Publishers, Košice, 2000. ISBN: 80-88922-13-5, in Slovak.

Comparison of software for modeling and simulation of robotic arms in the university environment

¹Peter NGUYEN (3rd year), ²Marek VACEK (1st year)
 Supervisor: ³Jaroslava ŽILKOVÁ

^{1,2,3}Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

¹marek.vacek@tuke.sk, ²peter.nguyen@tuke.sk, ³jaroslava.zilkova@tuke.sk

Abstract—This article is concerned with an evaluation and comparison of programs or toolboxes suitable for calculations and 3D modeling of industrial or experimental robotic arms. These are also appropriate as a teaching aid at universities.

Keywords— robotic arm, software, simulation, modeling, Robotics Toolbox, SpaceLib, RobotScene

I. INTRODUCTION

There are many programs used for robot modeling which facilitate solving of complex mathematical problems or 3D modeling. However, it is difficult to choose only one of them as the best. A lot of the programs are intended for industrial application and practice, e.g. programs designed by ABB, FANUC or KUKA. Nevertheless, there are programs intended for the use in the teaching process at universities.

This article focuses on such programs. All three of them have been developed by teachers and students from all around the world (Australia, Spain, Italy) for other academic communities. They are available for free download and use. The main characteristics of each of the programs, their pros and cons, and download and online help possibilities are discussed in this article.

II. DESCRIPTION OF SOFTWARE

A. Robotics Toolbox (EN)

Inventor: Peter I. Corke, Queensland University of Technology, Brisbane, Australia

System requirements – operating system: Microsoft Windows XP and higher. Required software: MathWorks Matlab 6 and higher – recommended version R2008

Robotics Toolbox is a toolbox for Matlab. It offers a lot of functions which are useful in the field of robotics, including kinematics, dynamics, and trajectory calculation. It can also be used for simulations, as well as for analysis of the results of experiments with real robots. The toolbox uses a general method for calculations for serial link manipulators. The user can create various types of robotic arms. There are examples

for robots which are already known, such as Puma 560 or Stanford arm. Robotics Toolbox also contains functions for manipulation and conversion between data types, such as vectors, homogenous transformations, and matrices, which are necessary for the expression of position and orientation in 3D.

The functions are described in a user-friendly way; however, they require a greater PC performance. If the user prefers effectiveness of calculations, it is possible to rewrite the functions into a more effective form, compile the m-file using Matlab, or create a MEX version (the increase of calculation speed is in some cases, e.g. dynamics, as much as 500x).

Robotics Toolbox is freely available at its homepage: <http://www.petercorke.com/robot>

The files are available either in the gzipped tar (.gz) format, or in the zip format (.zip). Information about the user (country, type of organization, and the intended use of the toolbox) is required at download.

After decompression of the files, the destination folder can be moved to any place at the HDD. Subsequently, the path to the folder has to be defined in Matlab:

File – Set Path – Add Folder...

The following message will show at each start of Matlab in the Command Window as a verification of a successful installation:

```
Robotics Toolbox for Matlab (release 8)
(c) Peter Corke 1992-2008 http://www.petercorke.com
installed in C:\Program
Files\MATLAB\R2008a\toolbox\robot
```

The toolbox should work with all versions of Matlab 6 and higher. However, Matlab's functions are frequently changed; therefore it is best to use at least the version R2007 or R2008. As Matlab versions 3.x and 4.x do not support objects, the toolbox will not function properly in them. On older version of the toolbox downloadable from Matlab's FTP server can be used, but it does not include some of the function of the current version.

The file robot.pdf (included in the compressed file with the program) is a user manual with a description of all functions

of the toolbox. A Google group called “Robotics Toolbox” has been created for any discussion concerning the program. It replaces all kinds of communication between users and the inventor. It can be found at:

<http://groups.google.com.au/group/robotics-tool-box>

An interactive demo is contained in the file `rtdemo.m`. This file is able to show most of the functions of the toolbox. After opening the file in Matlab, a GUI is shown, containing eight examples, which execute the required program in the Command Window.

Possibilities include transformations, trajectories, direct kinematics, inverse kinematics, jacobians, animations and visualization.

The following is an example of calculating the trajectories of the PUMA 560 robot, which will move the robot from its zero angle position to the upright (or READY) position. At first we can create a time vector, completing the motion in 2 seconds with a sample interval of 56ms.

```
t = [0:.056:2];
```

Then we compute a polynomial trajectory between the 2 positions using `jtraj()`

```
q = jtraj(qz, qr, t);
```

Since we know that for this particular trajectory most of the motion is done by joints 2 and 3, we can plot this process conveniently using standard MATLAB operations (Fig.1):

```
subplot(2,1,1)
plot(t,q(:,2))
title('Theta')
xlabel('Time (s)');
ylabel('Joint 2 (rad)')
subplot(2,1,2)
plot(t,q(:,3))
xlabel('Time (s)');
ylabel('Joint 3 (rad)')
```

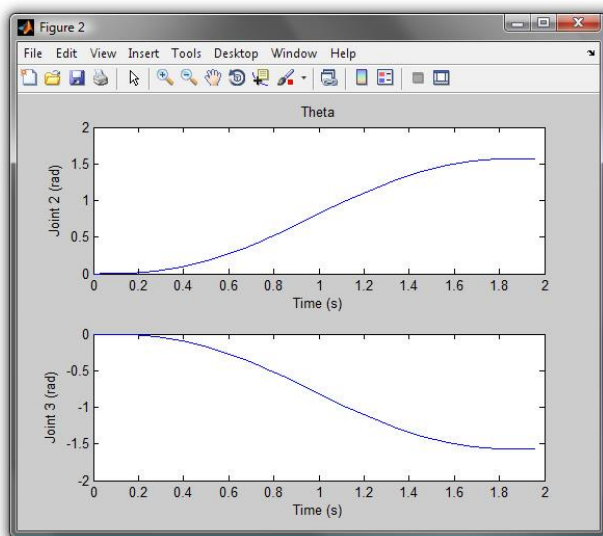


Fig. 1. Progress of rotation of joints 2 and 3 (Puma 560)

`Rtdemo.m` is also available in the Matlab help center.

Pros of the toolbox include a vast number of functions, great online support, and the practical examples in `rtdemo.m`.

Cons include the need for a Matlab license, inability of 3D robot modeling, and the change of commands and their syntax in the newer version.

B. SpaceLib (EN)

Inventor: Giovanni Legnani, Università Degli Studi di Brescia, Italy

System requirements – operating system: Microsoft Windows 2000 and higher. Required software: MathWorks Matlab 6 and higher

SpaceLib is a toolbox for Matlab. It is useful for the realization of programs for the analysis of kinematics and dynamics of mechanical systems. This toolbox is primarily used in robotics and biomechanics. It is intended for use as an aid while creating programs for the analysis of mechanical systems. This can be considered as a generalization of the transformation Denavit-Hartenberg matrices. The main advantage of such an approach is the creation of effective numerical programs.

Nowadays, three versions are available – two are designed for numerical simulations (in C/C++ and Matlab) and one for symbolic computation (Maple).

The Matlab version is particularly useful because of the possibility of fast development of computing programs. The C/C++ version is more suitable for achieving fast, highly effective numerical calculations. The Maple version's main use is for symbolic manipulations, but it can be used for the development of numerical programs as well.

The program includes functions for basic operations with matrices, vectors and planes, transformations of rotation matrices, matrices of acceleration and velocity, and Cardan-Euler angles.

The toolbox can be freely downloaded from the website:

<http://www.ing.unibs.it/~glegnani/>

or on the MathWorks website for file exchange, as well as on the Maple Application Center website.

The installation is identical with Robotics Toolbox. However, the verification can only be carried out using one of the functions.

The downloaded compressed file contains the source code in C/C++ and Matlab, a user manual in PDF, annotated samples of the program, and documents on the theory of solving the mathematical problems in robotics in PDF.

There is no online support.

Pros of the toolbox include a great number of functions and samples, a description of solving problems in robotics, and the inclusion of literature explaining the basic principles of robotics (in English).

Cons include no online support, and a need for a Matlab license.

C. RobotScene (ES)

Inventor: Antonio Romeo, Universidad de Zaragoza, Spain

System requirements – considering the use of 3-dimensional graphics, it is required that the computer on which RobotScene is to run fulfills the following minimum requirements: processor Intel Pentium 3 or higher. Operating system Microsoft Windows 98, 2000, Me or XP. OpenGL compatible video card.

RobotScene is a specific program which provides graphical interpretation of all concepts and industrial robot design procedures (creation of segments, joints, systems of coordinates, use of Denavit-Hartenberg parameters, etc.). Moreover, it enables programming of previously created robots, thus enabling modeling of robotic arm segments, creation of a robot (Fig.2), creation of the environment, and the final programming of a robot.

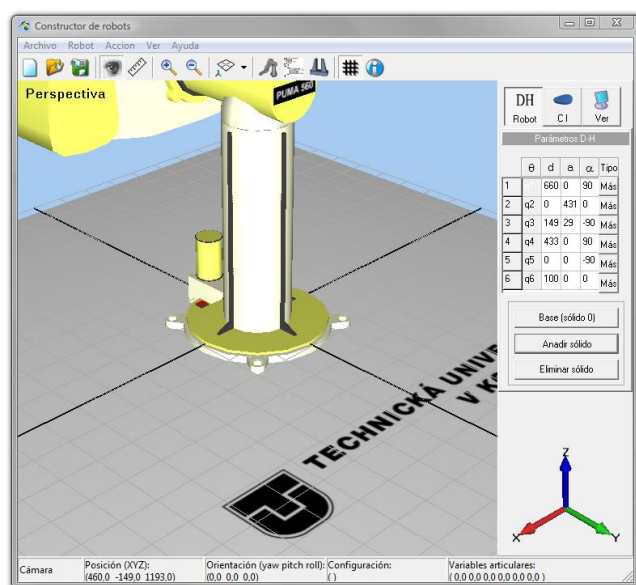


Fig. 2. GUI for the creation/construction of a robot

There are a lot of other platforms which can be used in robotics, and which provide simulation tools (e.g. Mellado 2003). Such programs are not, however, focused on modeling and design of robotic arms. The aforementioned platforms (Robotics Toolbox or SpaceLib) provide tools useful for the calculations of kinematics and dynamics of a robot, but do not include a GUI, and do not provide tools to program the robot's movement.

The program is freely downloadable from the website of the University of Zaragoza:

<https://moodle.unizar.es/mod/resource/view.php?id=17880>

After decompression of the file, the user should open the resulting folder and execute the file RobotScene_setup.exe to start the installation. The installer uses the common interface with an installation wizard which is helpful regarding any possibilities and parts of the installation process, when the choice of a folder into which the application installs is required.

The user manual is available at the website of the University of Zaragoza. However, it is only available in Spanish. The inventor has confirmed that an English version is not planned.

Online support is provided at romeo@unizar.es.

This program contains models of robot parts as well as functional models of robots by companies such as FANUC, MITSUBISHI, POLAR, etc. However, these are not sufficient for showing the operation and presentation of the program.

Pros of the program include simple and easily understandable GUI, no need to use other software to make the program functional, possibility of 3D modeling of robot parts and the subsequent assembly of a robot, and great online support. Moreover, the program takes up very little space on HDD after installation.

Cons include the need to use other software to carry out calculations of mathematical tasks, frequent outages of the website from which the program and the manual can be downloaded, and the fact that both the program, and the manual are only available in Spanish.

III. CONCLUSION

It can be noticed that even though the first two programs focus on similar topics – the solution of mathematical problems in robotics – there is a different approach to the explanation of the processes to the user. Compared to Robotics Toolbox, SpaceLib has better explanations of the topic and mathematical solutions of tasks in robotics. However, Robotics Toolbox is updated more frequently, and therefore exploits the potential of the new Matlab versions better.

RobotScene differs from the first two programs by its focus on 3D modeling and programming of the movement of robotic arms. When used in combination with previous calculations in Matlab, it is a valuable tool for further modeling of a robot, including its simulations in a virtual environment.

Robotics Toolbox is the most suitable for application in the research area. It supports kinematics calculations, dynamics calculations and trajectories generation directly in connection with other Matlab and Simulink functions, and therefore can be used for complete support and fine-tuning of the simulations of real systems. SpaceLib and RobotScene are not suitable for use in the research area because of poor updates (SpaceLib) or no real system output (RobotScene).

REFERENCES

- [1] CORKE, Peter I.: A Robotics Toolbox for MATLAB. CSIRO Division of Manufacturing Technology, 2008, [cit. 2010-7-14]. PDF.
- [2] LEGNANI, G. – ADAMINI, R. – ZAPPA, B.: SpaceLib in Matlab (Version 2.2). Università di Brescia, 2005, [cit. 2010-11-14]. PDF.
- [3] ROMEO, A.: Software de creación y explotación de escenarios robóticos. Universidad de Zaragoza, 2009, [cit. 2010-11-14]. PDF.
- [4] VACEK, M.: Modelovanie priemyselných robotov. Master's thesis. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2011.

Constructional solution of sensors for measuring of intra-abdominal pressure

¹Michal JURČIŠIN (3rd year), ²Pavol CABÚK
Supervisor: ³Stanislav SLOSARČÍK

^{1, 2, 3}Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic

¹michal.jurcisin@tuke.sk, ²pavol.cabuk@tuke.sk, ³stanislav.slosarcik@tuke.sk

Abstract— The article deals with the design of pressure sensors based on thick film technology and thin film technology. Suitable type of designed pressure sensor should be used in developed automation system for measurement of intra-abdominal pressure by noninvasive method. First part of paper deals with design, test measurements and properties evaluation of sensor based on thick film technology. Second part of paper mentioned alternative solution based on thin film technology applied in design of capacitive pressure sensors.

Keywords—capacitive sensor, intra-abdominal pressure, LTCC technology, pressure sensor.

I. INTRODUCTION

Abdominal compartment syndrome (AbCS) is caused by the increase of intra-abdominal pressure (IAP) from tissue tumescence or from accumulation of free liquid in abdominal cavity. AbCS is a 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 precise method of the IAP measuring is applying pressure sensor through abdominal wall directly into abdomen cavity. Measuring of pressure directly in abdomen cavity via a 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 the fact that urinary bladder works like a passive transmitter IAP on internal water filling. Empty urinary bladder is filled by 20-50 ml of sterile saline via catheter. [2] Pressure in abdomen cavity is transferred across the bladder on filled solution and so on as 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 emptying of all system. For measurement of IAP commercial systems are used. An advantage of commercial systems is that they are delivered in assembled state a disadvantage is the necessity of manual operation after each cycle.

II. CONCEPT OF AUTOMATION MEASURING SYSTEM OF INTRA-ABDOMINAL PRESSURE

For using of developed automation measurement system in clinical praxis it is necessary to state basic requirements. Due to application field it is necessary to ensure sterility of the

system during measurement. Distribution of intravesical pressures grades used in clinical practice is listed in Table I. In the second grade i.e. at a pressure of 2,66 kPa is necessary to decompress the abdominal cavity what is reached by cutting the abdominal wall and keep it open. Measurement of IAP is carried out at one hour intervals with the following procedure:

- system preparation for measurement - prefilling of bladder and system tubes with exact volume of saline,
- realization of measurement in the time interval of several minutes,
- drying of system.

TABLE I
DISTRIBUTION OF PRESSURE GRADES USED IN CLINICAL PRACTICE

Grade	cmH ₂ O	p [kPa]
1.	16 – 20	1,59 – 1,99
2.	21 – 27	2,13 – 2,66
3.	28 – 34	2,79 – 3,33
4.	> 35	> 3,33

III. THICK FILM PRESSURE SENSOR

A. Design of thick film pressure sensor

One of possible options was realization of pressure sensor based on thick film (TF) technology. On circle shaped LTCC was realized electrical circuit in connection to Wheatstone bridge by application of screen printing technology. Scheme and topology of realized circuit is shown on figure 1. Geometrical dimensions of membrane were designed with respect to highest sensitivity of sensor and also to achieve the smallest volume of cavity under the membrane.

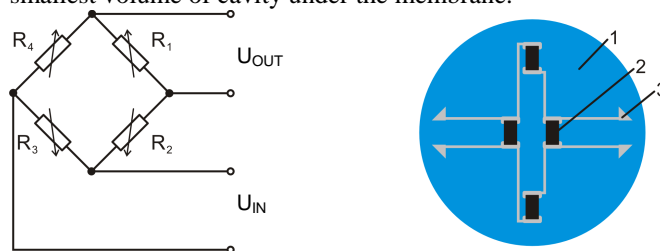


Fig. 1. Scheme and resistors topology realized on LTCC membrane ($R_1 - R_4$ – Wheatstone bridge resistors) 1 – LTCC membrane, \varnothing 37 mm, thickness 220 μ m after burning; 2 – TF resistor $R = 100$ k Ω/\square , $\alpha_{TKR} = 100$ ppm/ $^{\circ}$ C; 3 – output for connection of wires [33]

B. Construction of thick film pressure sensor

Realized TF pressure sensor was applied into package according Fig. 2.

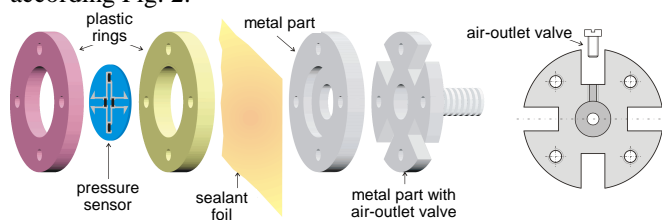


Fig. 2. Parts of package with mounted pressure sensor

Package consists of plastic rings for fixation of pressure sensor (LTCC membrane) and metal parts for creation of proper cavity under the pressure sensor. Diameter of cylindrical cavity has to be sufficient for correct measurement and also respect volume of solution filled into bladder. For isolation of direct contact of solution with pressure sensor was used sealant foil inserted between plastic ring and metal part. Isolation of every single part of package was used isolation silicon paste. All parts was assembled and fixed by screws. Because the measurement system use a pressure of liquid column, it was necessary to use an air-outlet valve to ensure continual liquid column without air bubbles in reconnection tubes and sensor package.

C. Testing of thick film pressure sensor

Principle of testing of pressure sensor realized by thick film technology intended for measurement of intra-abdominal pressure is shown on Fig. 3. Measuring system consist of package with pressure sensor, throttle valves (V1, V2 and V3) syringe with liquid solution and balloon (imitation of patients bladder). For supply of pressure sensor was used stabilized voltage source. Output voltage was measured by digital multimeter.

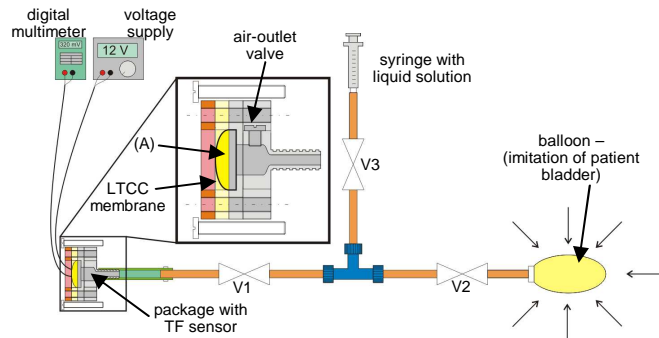


Fig. 3. Principal scheme of measurement with thick film pressure sensor

At the beginning of the measurement was opened valves V1, V3 and closed valve V2. In this step was prefilled cavity (Fig. 3 (A)) under LTCC membrane by syringe. At the same time was opened air-outlet valve. After venting was air-outlet valve and valve V1 closed. In next step was opened valve V2 and balloon was filled by syringe. Finally was closed valve V3 and valves V1 and V2 opened. This step starts test measurement.

During measurement was pressure increased in interval 0 – 6 kPa, what caused voltage changes of Wheatstone bridge. Measured voltage increased from value 295 mV at unloaded state to 342 mV at fully loaded state. The voltage change was 47 mV. Voltage changes depended on pressure changes is shown on Fig. 4.

$$U = f(p)$$

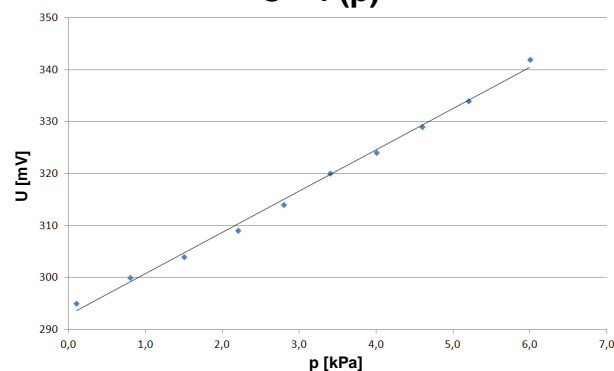


Fig. 4. Output voltage course as function of pressure of thick film pressure sensor

D. Properties evaluation of thick film pressure sensor

Measurement system with TF pressure sensor cannot be considered as sterile on the ground of application of silicon paste, which purpose is isolation of single parts of package with sensor. Measuring system with TF pressure sensor allows measurement realization and initialization to starting point in time interval shorter as required by clinical praxis. Other disadvantage of measuring system is necessity of venting, which realization is problematic. These facts determine application of TF pressure sensor as inconvenient in consider of practical realization of automated measuring system.

IV. CAPACITIVE PRESSURE SENSOR

Other solution is realization of sensor using thin film technology. 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. 5. As dielectric of capacitive sensor was used saline (0,9% NaCl aqueous solution).

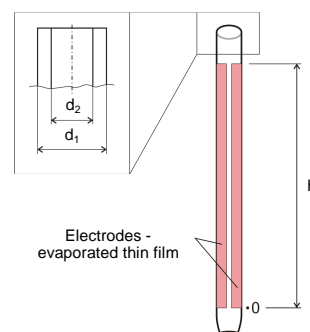


Fig. 5. Capacitive pressure sensor (d_1 – outer tube diameter, d_2 – inner tube diameter, h – height of evaporated electrode)

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.

In Table II parameters of used glass tubes are listed. 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 according to Fig. 6.

TABLE II
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

Calibration of capacitive sensors was realized in temperature stabilized conditions at 20°C. Control measurements with temperature changes in range from 20°C to 50°C does not prove temperature dependence of electrical parameters on capacitive pressure sensor.

$$C = f(p)$$

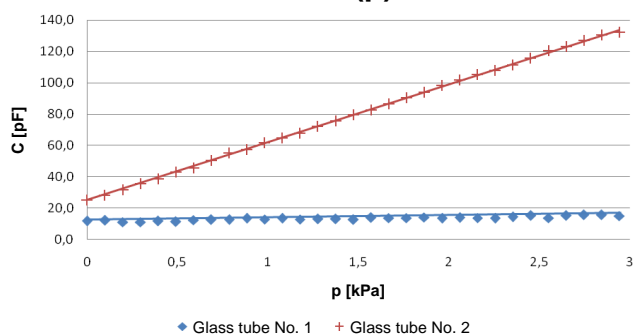


Fig. 6. Calibration curves of tested glass tubes.

V. CONCLUSION

The use of capacitive sensors in concept of automated measuring system is evaluated as better solution in compare with thick film pressure sensors. Capacitive 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 the diameter of glass tube No. 2. Using tubes with larger diameters cause increase of internal volume of the tube making them unsuitable for the application with limited volume of saline filled into the bladder.

Whole system will simplify measurement process in focus of medical staff and eliminate the human mistakes. In the first step was solved sterility of system, measurement via capacitive sensor. Automated measuring system will be tested in laboratory conditions. (Project VEGA 1/0059/12)

REFERENCES

- [1] T. Molčányi, J. Magdo, R. Raši, T. Cestická, J. Michlík, A. Molčányiová, "Možnosti ošetrovania laparostomie při abdominálnom kompartment syndrome" in *Novinky v anesteziológii, intenzívnej medicíne 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] I.L. Kron, P.K. Harman, S.P. Nolan, "The measurement of intra-abdominal pressure as a criterion for abdominal re-exploration".
- [4] M. Jurčišin, S. Slosarčík, P. Cabúk, "Automation measurement of intra-abdominal pressure", In: *Advanced Numerical Modelling*: 28. - 31.8.2011, Zielona Góra. – Warsaw: Electrotechnical Institute, 2011, p. 95-96., ISBN 978-83-61956-02-0

Design of Microstrip Hairpin Band Pass Filter for UWB Application

¹Kornel RUMAN (1st year), ²Ondrej KOVÁČ (1st year), ³Igor VEHEC
Supervisor: ⁴Alena PIETRIKOVÁ

^{1,3,4}Dept. of Technologies in Electronics, KTE TU of Košice, Slovak Republic

²Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

¹kornel.ruman@tuke.sk, ²ondrej.kovac@tuke.sk, ³i.vehec@tuke.sk, ⁴alena.pietrikova@tuke.sk

Abstract—This paper deals with design and simulation of a microstrip hairpin band pass filter for UWB (Ultra Wide-Band) radar applications. This paper demonstrates the design of five-section microstrip hairpin band pass filter using the *Filter Solution 2011* from Nuhertz and simulation by *Sonnet Lite 13.55* from Sonnet Software. Simulated results of pass band response, return loss response and current density of microstrip hairpin band pass filter for UWB radar are presented. It assesses the suitability of hydrocarbon ceramic laminate (Rogers RO4003C) substrate for the production of hairpin microstrip band pass filter for high frequency area. It examines potential possibility of manufacturing of such a filter based on LTCC technology. The presented filter is an example of a microstrip hairpin band pass filter for particular laboratory UWB radar.

Keywords— microstrip hairpin Band Pass Filter, UWB radar, LTCC, hydrocarbon ceramic laminate (Rogers RO4003C).

I. INTRODUCTION

The hairpin filter is one of the most popular microwave frequency filters because it is compact and does not require grounding. Its form is derived from the edge-coupled resonator filter by folding back the ends of the resonators into a “U” shape. This reduces the length and improves the aspect ratio of the microstrip significantly as compared to that of the edge-coupled configuration [1, 2].

UWB radar systems are used in a variety of applications. UWB are well-known for usefulness in localization and radar applications. The use of UWB signals can provide distance measurements with high accuracy. However, big challenge is design, simulation and construction of a quality filter for removing unwanted frequencies. In particular, design and simulation of classic microstrip hairpin band pass filters for UWB radar applications with a bandwidth of 6 to 8 GHz, is the challenging problem [3]. The Bandwidth of the filter was selected, seeing that ECC (Electronic Communications Committee) in Europe set allowed frequency band from 6 to 8.5 GHz for UWB radar applications [4].

Nowadays, there are a variety of programs for simulation and design of filters for high frequency area, at the market. When the filters are designing for high frequency area, there's necessary an overall response simulation. The hairpin filter was designed in software *Filter Solution 2011* from Nuhertz, the worldwide leader in low and high frequency filter synthesis and analysis. The design was exported to *Sonnet Lite*

13.55 from Sonnet Software and simulated with using planar EM (Electromagnetic) analysis [5, 6].

The substrate has a major impact on the quality and stability of the filter. There are many substrates with various dielectric constants that are used in wireless applications. Therefore not every substrate is suitable for high frequency area. At the market there is a large amount of available substrates with different parameters, so choosing the appropriate substrates should pay close attention. Those with high dielectric constants are more suitable for lower frequency applications. The FR4 laminate is the most common electronic carrier for circuits operating below the microwave frequency. Beyond 1 GHz, the laminate properties (like ϵ_r , $\tan \delta$, roughness, etc...) of the FR4 become nonlinear. It becomes difficult to precisely characterize the FR4 at the desired microwave frequency. Beyond 3 GHz, the use of the FR4 is not recommended anymore, in case of unacceptable attenuation. This is the reason why we chose a substrate based on hydrocarbon ceramic laminate (ROGERSR O4003C) [7].

II. FILTER DESIGN

A. Substrate

For correct design and simulation is necessary to know the parameters of the substrate. As substrate we chose hydrocarbon ceramic laminate (ROGERS RO4003C). Parameters of used substrate based on hydrocarbon ceramic laminate (ROGERS RO4003C) are in Table 1.

TABLE I
ROGERS RO4003C SUBSTRATE SPECIFICATION

Property	Symbol	Typical Value (Units)
Dielectric Constant	ϵ_r	3,55
Dissipation Factor Tan	δ	0,0027
Dielectric Height	h	0,008" (0,203 mm)
Resistivity Compared to Copper	ρ	1
Metal Thickness	M	½ oz. (17µm)
Volume Resistivity	-	1.7×10^{10} (MΩ•cm)
Surface Resistivity	-	4.2×10^9 (MΩ•cm)
Moisture Conductivity	-	0.71 (%)
Melting temperature Tg	-	>280 (°C DSC)

Low dielectric loss allows hydrocarbon ceramic laminate material to be used in many applications where at higher

operating frequency are the conventional laminates circuit boards limited.

The temperature coefficient of dielectric constant is among the lowest of any circuit board material, and the dielectric constant is stable over a broad frequency range (Figure 1). This makes it an ideal substrate for broadband applications. Hydrocarbon ceramic laminate has also melting temperature T_g of $>280^\circ\text{C}$ (536°F) so its expansion characteristics remain stable over the entire range of circuit processing temperatures [8].

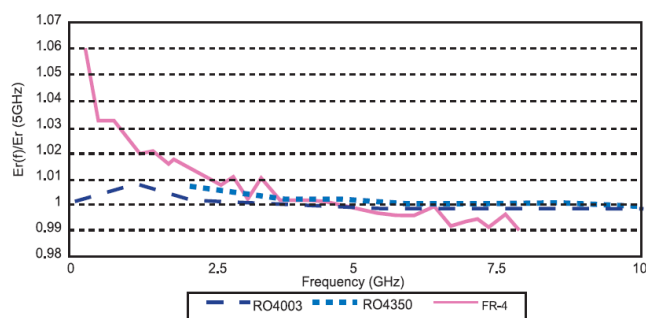


Fig. 1. Change in dielectric constant due to the increasing frequency [8].

From the Figure 1 it can be concluded that the best from comparison substrate is hydrocarbon ceramic laminate RO4003C. Therefore, all presented results of the simulation are for hydrocarbon ceramic laminate substrate (Rogers RO4003C).

B. Initial Design

This filter was designed for a flat response over the 6 GHz to 8 GHz band. After simulation, we decided to choose the filter type Chebyshev I and distrib implementation, which require attenuation and impulse response. A five-section microstrip hairpin filter topology was chosen to provide a band pass response centered at 7GHz with a bandwidth of approximately 2 GHz.

The pass band magnitude of Chebyshev I and Elliptic filters oscillate up and down within a user-specified boundary. The higher order of the filter, the more oscillations occur. These oscillations are known as Pass Band Ripple [5]. The pass band ripple of 0.05 dB is used. The width of the hairpin lines was set at 200 μm and width between the hairpin legs (hairpin gaps with) was set to 2 mm. The filter design dimensions were optimized to meet the specifications in the pass band. We also exploit function optimization, provided by *Filter Solution 2011*, to our filter design. It takes much more computing time, but attempts to provide a more accurate solution than just the raw synthesis. Resulting layout of the hairpin filter is shown in the Figure 2.

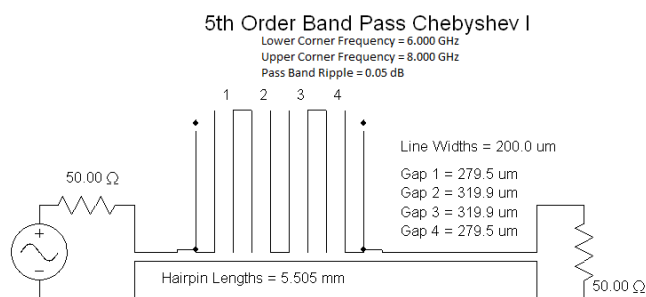


Fig. 2. Layout of the five-section microstrip hairpin filter, designed with the help of Filter solution 2011.

The area allocated by the filter is approximately 13.5 by 5 mm (0.531 x 0.197 in). The microstrip hairpin band pass filter input and output are matched to 50 ohm characteristic impedance. The 50 ohms choice is a compromise between power handling capability and signal loss per unit length, for air dielectric [9].

III. RESULTS AND ANALYSES

The design was exported to *Sonnet Lite 13.55* from Sonnet Software. *Sonnet Lite* is a free feature-limited version of Sonnet's professional Sonnet Suites, which provides EM analysis for thousands of companies all over the world. Many major manufacturers of high-frequency components and boards depend on Sonnet Software for analysis of their predominantly planar high-frequency designs from 1 MHz through several THz. The exported 3D diagram of the hairpin filter is shown in the Figure 3.

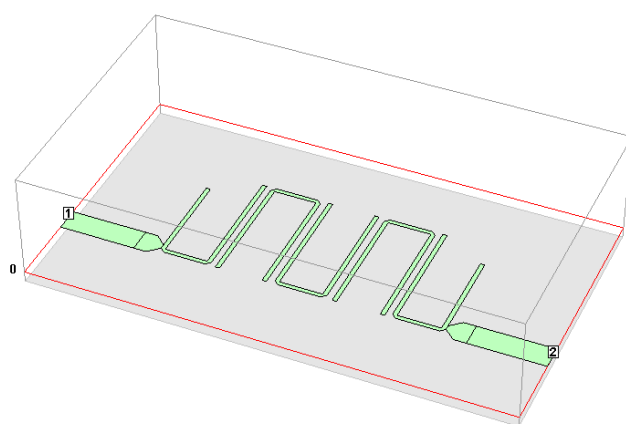


Fig. 3. 3D layout of hairpin filter

This layout data were used for simulation in Sonnet Lite 13.55 with using planar EM (Electromagnetic) analysis. Pass band results of the EM simulation is shown in the Figure 3.

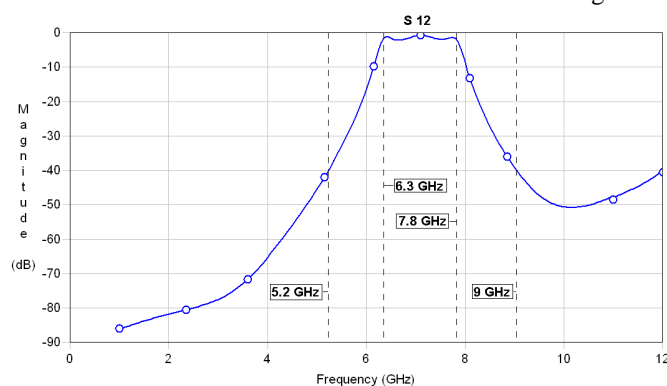
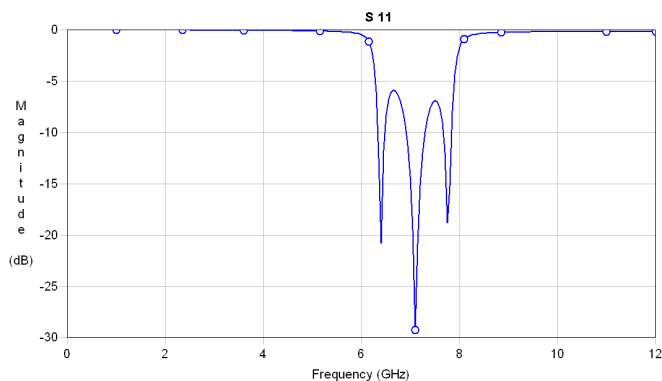


Fig. 4. Pass band response of the 5th order hairpin filter

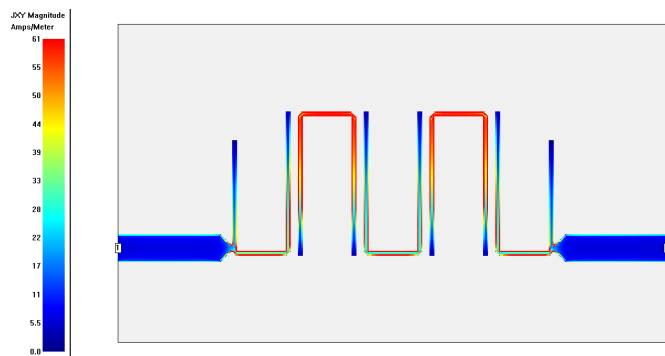
EM simulation was purposely made only up to 12 GHz. The Figure 4 shows that pass band is a bit narrower. The corner frequency is shifted from 6 to 6.3 GHz and from 8 to 8.7 GHz. Figure 4 shows that designed hairpin band pass filter meets the minimum attenuation of -40 dB with the corner frequency shift of 1 GHz in suppress band.

A return loss result of the EM simulation is shown in the Figure 5.

Fig. 5. Return loss response of the 5th order hairpin filter

EM analysis results from *Sonnet Lite 13.55* indicates that the pass band and return loss response meets the design criteria.

We use current density viewer to view results from an EM analysis as a colour shaded plot (Figure 6). The colours can represent either the magnitude of current density or charge. By displaying a sequence of current or charge density plots, each one at a slightly different frequency or phase, the current density viewer is able to make a real time animation. The plots shown with the current density viewer give new insight into the workings of circuit by showing the "hot spots" and "cold spots" of circuit.

Fig. 6. Current density of the 5th order hairpin filter

IV. CONCLUSION

The 5th order hairpin band pass filter was designed and simulated for band from 6 to 8 GHz. The layout of hairpin band pass filter was created by the Filter solution 2011 and it was used for simulation of filter in *Sonnet Lite 13.55* with planar EM analysis. The pass band and return loss were presented. Hairpin filter will be fabricated on hydrocarbon ceramic laminate substrate (ROGERS RO4003C) and measure by Signal Analyzer N9020A.

ACKNOWLEDGMENT



This paper was developed with support of the project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055, that is co-financed from Structural Funds EU ERDF within Operational programme Research and Development OPVaV-2009/2.1/03-SORO and preferred axis 2 Support of Research and Development.

REFERENCES

- [1] Jia-Sheng Hong, "Microstrip Filters for RF/Microwave Applications," John Wiley & Sons, Inc., Hoboken, 2nd edition, New Jersey, 2011, 655.
- [2] Jia-Sheng Hong and M. J. Lancaster, "Microstrip Filters for RF/Microwave Applications," John Wiley & Sons, Inc., Hoboken, New York, ISBN 0-471-22161-9, 2001, 471.
- [3] I. J. Immoreev, and J. D. Taylor, "Future of radars," Ultra Wideband Systems and Technologies, 2002. pp. 197-199.
- [4] <http://www.cept.org/ecc>
- [5] <http://www.nuhertz.com/profile.html>.
- [6] www.sonnetusa.com.
- [7] Nikholas G. Toledo, "Practical Techniques for Designing Microstrip Tapped Hairpin Resonator Filters on FR4 Laminates", in Communications Engineering Division, Advanced Science and Technology Institute, Diliman Quezon City, Philippines 1101
- [8] <http://microboone-docdb.fnal.gov/cgi-bin/RetrieveFile?docid=1896;filename=RO4000%20%20Laminates%20-%20Data%20sheet%20.pdf;version=1>.
- [9] <http://www.microwaves101.com/encyclopedia/why50ohms.cfm>.

Design of the I-Q Demodulator for UWB Applications

¹Martin LIPTAJ (3rd year PhD.), ²Matej ŽIGA (2nd year Ing.)

Supervisor: ³Pavol GALAJDA

^{1,2,3}Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

¹martin.liptaj@tuke.sk, ²matej.ziga@student.tuke.sk, ³pavol.galajda@tuke.sk

Abstract—In the article the initial design problems and results of the HF I-Q demodulator design which is a part of I-Q Extension kit for UWB sensor system are presented. The I-Q demodulator is designed in order to match the Electronic Communication Committee frequency band intended for UWB applications from 6 GHz to 8 GHz.

Keywords—I-Q demodulator, Mixer, Sensor system, SiGe BiCMOS, Ultra-WideBand (UWB), Electronic Communication Committee (ECC)

I. INTRODUCTION

The rapidly growing field of UWB (Ultra-Wideband) applications in various areas of present human existence such as positioning and localization tasks or in communication services as well as in sensor implementations pushes the requests for a new enhanced UWB radar systems. One of the very promising solution for the UWB device realization is based on so called M-sequence approach, where the operation is based on a special type of the M-Sequence presented in [1]. This UWB sensor system is simply expandable with I-Q demodulator on side of receiver [2]. I-Q demodulation is very useful if we need to obtain both the magnitude and phase of the received signal. The remarkable approach with an I-Q demodulator is an easy and quick record all the information by just measuring two voltages. Therefore, I-Q demodulator is an important building block in RF receiver with digital modulated baseband signals.

In this paper a design of I-Q demodulator for use as extension with an UWB M-Sequence sensor system is presented. I-Q demodulator is designed for operation in the range from 6GHz to 8GHz in order to match the Electronic Communication Committee (ECC) frequency band intended for UWB applications [3]. The demodulator contains the chip with the wide-band mixers pair, the delay line for 90° phase shifting for 7 GHz LO signal, band-pass and low-pass filters. The signal from this I-Q demodulator will be digitized by flash A/D converter and then processed in further digital signal processing unit in sensor system. Other common known alternatives for signal acquisitions/sampling, e.g. Sigma-Delta approach, are recently not considered due to realization challenges because of handling of ultra-wideband signals in ECC, or FCC licensed ranges with required equivalent sampling rates of over 20GHz and also later intended implementations in cm- and mm-wave ranges (around 24GHz and 60GHz) with realized I-Q stage as interstage, as well as due to compatibility to the mother sensing unit from TU Ilmenau [1].

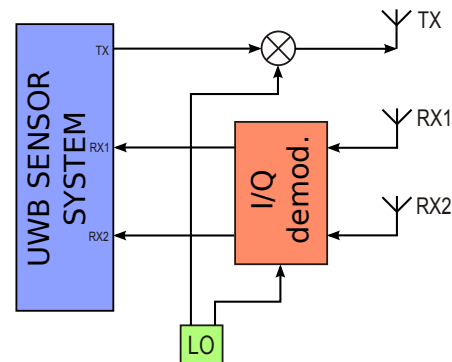


Fig. 1. Block diagram of sensor system with extension kit

The presented paper consists of a three parts. In the first part a basic theoretical principles of I-Q demodulator are described. The second part deals with the presented I-Q demodulator design, especially with the mixers, delay line and power divider. Finally, a PCB design and several simulations of I-Q modulator system are described.

II. BASIC THEORY OF I-Q DEMODULATOR

The analog I-Q demodulator uses two matched demodulator circuits to convert the RF input signal directly to baseband analog I and Q signals. These signals are subsequently converted to digital data using an A/D converters. The functionality of this circuit is based upon the RF input signal being split and mixed in two wideband mixers with two local oscillator (LO) signals that have a 90° phase shift between them. This 90° phase shift provides the mechanism to distinguish the I and Q components of the RF signal. The mixer outputs are low-pass filtered to remove the high-frequency mixing products, providing baseband analog I and Q signals that are sampled and converted to digital values. The parallel nature of the analog I-Q demodulator requires the two ports to be very closely matched to each other for accurate I-Q measurements. Also, the quadratic phase shift must be exactly 90° at all frequencies. This nature of the conventional analog I-Q detector makes it susceptible to errors associated with gain matching, DC offsets, quadratic phase errors, carrier leakage, and impedance matching. These errors can be difficult to completely eliminate or compensate for, causing RF measurement errors [4].

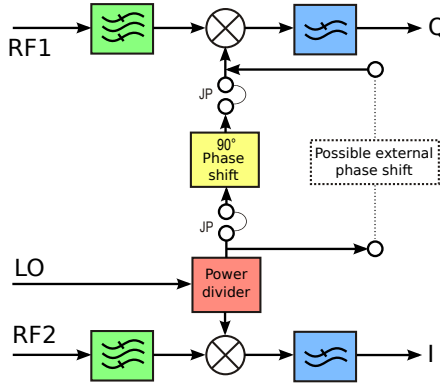


Fig. 2. Block diagram of proposed I-Q demodulator

III. DESIGN

Despite of above mentioned disadvantages (and with reflection, that these disadvantages are possibly rejected with a digital I-Q demodulator), the design of the analog I-Q demodulator is presented. For implementation as UWB sensing system frontend is required linear dynamic range of better 50 dB. Recent I-Q demodulator will be a part of complex system, whose block diagram is shown on Fig.1. It consists of M-sequence sensor system, simple up-converter and recent I-Q demodulator. Simple up-converter is connected to transmitter output from sensor system. Then the mixed signal is transmitted by the transmitter antenna. Receiver antennas are connected to inputs of recent I-Q demodulator, whose outputs are connected to receiver inputs of M-sequence sensor system [2]. Because of working principle of mother UWB sensing unit [1] the unwanted signal artifacts coming due to linear errors can be calibrated up to certain degree.

It has to be pointed out that despite well-known theoretical IQ-demodulator structure, due to the handling of real UWB signals on the carrier (PCB) level the housing parasitic have to be understood. This is by contrast with narrow band systems challenging and in praxis more or less successive process. On the other hand, on-wafer measurements, which are often presented in publications, are useful for pre-selection of the ICs but not at all for the final characterization and implementation of a circuit. The reason of this is that in practice only mounted chips can be used. Therefore, the packaging issues have to be considered during the unit design flow too.

I-Q demodulator, whose block diagram is shown on Fig. 2., consist from input band-pass filters, single chip of wide-band mixer pair, delay line for 90° phase shift and output low-pass filters. They will be described latter in this article.

A. Power divider

In electronic circuits is often necessary, even desirable, to divide the various signals. This dividing or distribution of the signals can be provided by power dividers. Power divider in recent I-Q demodulator is used to divide a signal from oscillator between LO input of I mixer and signal for 90° phase shifting delay line. Because of the simple design and small size, the resistive power divider was selected [5]. The power divider consists of three resistors (Fig. 3). Resistor values for a N-way power divider is possible calculated from equation

$$R = Z_0((N - 1)/(N + 1)) \quad (1)$$

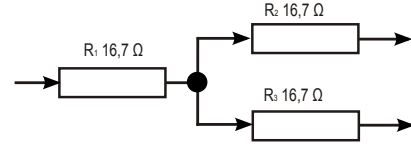


Fig. 3. Connection of power divider

where Z_0 is a characteristic impedance of the system and N is a number divided signal ways. For 2-way divider ($N = 2$), the resulting resistance is approximately equal to 16.7Ω.

B. Phase shifter

Design of the phase shifter can be realized in several methods. One of them is a design as the nonuniform transition line, which strip width is expanded in a truncated Fourier series. Then, the optimum values of the coefficients of the series are obtained through an optimization approach to have low phase shift error and low reflection coefficient in desired frequency bandwidth [6]. Because this calculation is relatively complex and realization is not simple, this method was not used for design of phase shifter. Delay line described in [7] was used to obtain 90° phase shifting for specific frequency. This delay line is formed by simple micro-strip line on printed circuit board (PCB). Its length (Δl) is given by equation

$$\Delta l = \frac{\Delta \Theta}{\beta} \quad (2)$$

where $\Delta \Theta$ is phase shift in radians and β is phase constant defined as

$$\beta = \frac{2\pi}{\lambda_G} \quad (3)$$

where λ_G is a wave length of the substrate. The substrate wave length is depended on relative permittivity ϵ_r of substrate and frequency of shifted signal and can be described by equation

$$\lambda_G = \frac{\lambda_{air}}{\sqrt{\epsilon_r}} = \frac{c}{f\sqrt{\epsilon_r}} \quad (4)$$

where c is a speed of light, f is the frequency of shifted signal and ϵ_r is relative permittivity of substrate. It means that with increasing frequency of shifted signal, the length of delay-line decreases.

For a delay line a enhanced Rogers 4003c carrier PCB was selected, which has a convenient RF parameters (especially relative permittivity $\epsilon_r=3.55$, thermal stability, etc.) [8]. For 90° phase shifting, which means that $\Delta \Theta$ approximately equals 1.57 rad and frequency of LO signal equals 7 GHz, a calculated length of delay line equals 0.005687 m.

The simulation result of phase shifting stripline is shown in Fig 4. As can be seen, the simulated shift of phase is equal 90°@7 GHz.

C. Mixers

Mixers in I-Q demodulator are most important devices. Recent I-Q demodulator contains a monolithic mixers chip designed and fabricated in 0.35μm SiGe BiCMOS technology and packaged in QFN32 package. A chip die consists of two identical active wide-band mixers and one differential amplifier. The mixers are based on Gilbert-cell topology. Their cut-off mixed frequency is up to 10 GHz and on chip isolation between LO - IF and RF - IF ports is better than 50 dB

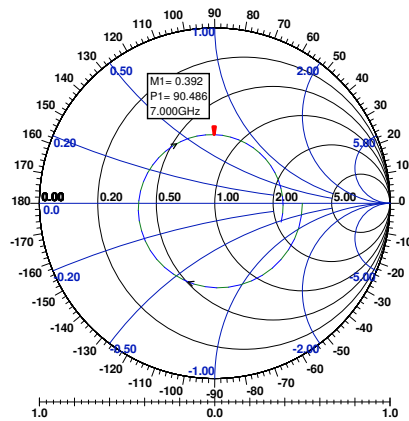


Fig. 4. Simulated phase shift result in Schmitt chart

respectively. In differential mode transient gain between IF-RF ports is around 6 dB, provided by 150 mV DC bias voltage on LO input.

D. Filters

The main aim of the filter is to attenuate the unwanted frequency components which appear in the I-Q demodulator spectrum. Sensor system is designed in order to match the ECC frequency bandpass. Therefore the received signals on I-Q demodulator inputs are necessary selected with band-pass filters. For this purpose recent I-Q demodulator was selected commercially fabricated band-pass LTCC filters BFCN-7900+ from Mini-Circuits company [9]. After mix-down conversion in I-Q demodulator is also necessary rejected the high frequency signals, such as LO and other signals, which can be considered as unwanted noise signals etc. Mentioned UWB sensor system operate in frequency band from DC to 2 GHz, that means the low-pass aliasing filter is necessary. For this purpose commercially fabricated low-pass LTCC filter LFCN-3400+ also from Mini-Circuits company was selected [10].

IV. REALISATION

As mentioned above we select enhanced Rogers 4003C carrier PCB for is excelent RF properties. The prototype PCB

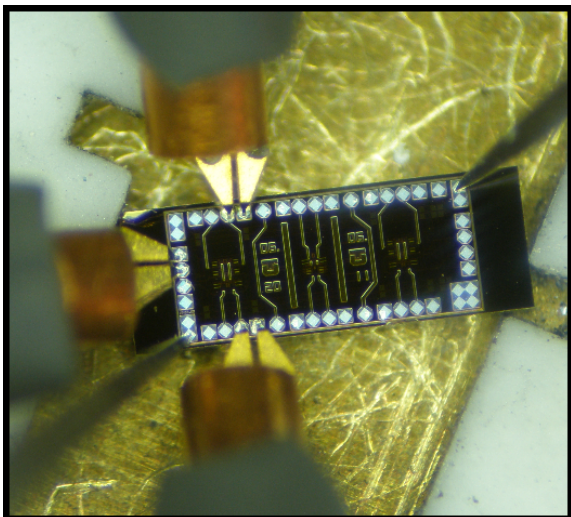


Fig. 5. Mixers chip measured on wafer probe station

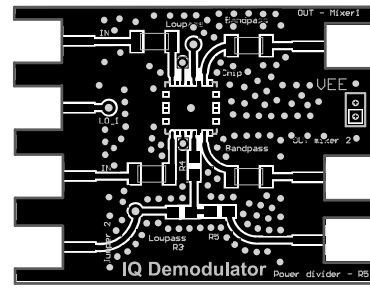


Fig. 6. The designed PCB for I-Q modulator - the top layer

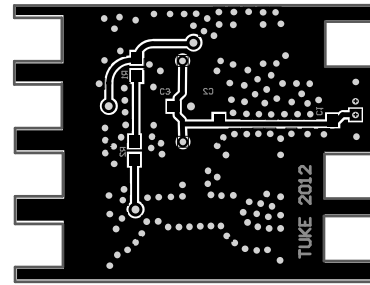


Fig. 7. The designed PCB for I-Q modulator - the bottom layer

is double - layers. The layout of the top and bottom layers of PCB is shown in Fig. 6 and Fig. 7, respectively. By default the delay line remains unconnected. Its connection is possible by soldering two of the pads together. For better variability of the delay it is possible to pass the PCB delay by connecting the external delay.

V. CONCLUSION

The initial design of the high frequency I-Q demodulator design a part of I-Q Extension kit for UWB sensor system is presented. In the future work, immediately after the manufacturing of I-Q demodulator, it will be important to measure and characterize it.

ACKNOWLEDGMENT

The authors will acknowledge Martin Kmec and also the whole team at EMT, TU Ilmenau for support. This work is also the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

- [1] J. Sachs and P. Peyerl, "A New Principle for Sensor-Array-Application," *Proceedings of 16th IEEE Instrumentation and Measurement Technology Conference*, pp. 1390-1295, 1999, IMTC/99 Venice, Italy.
- [2] J. Sachs, M. Kmec, R. Zetik, P. Peyerl, and P. Rauschenbach, "Ultra wideband radar assembly kit," *Geoscience and Remote Sensing Symposium*, 2005, iGARSS '05. Proceedings. 2005 IEEE International.
- [3] TUSR, "General permit No. VPR - 02/2011," *Available on internet*, 10.3.2012, <http://www.teleoff.gov.sk/data/files/22711.pdf>.
- [4] C. Ziomek and P. Corredoura, "Digital I/Q Demodulator," *Particle Accelerator Conference, Proceedings of the 1995*, pp. 2663 - 2665, 1995, vol.4, Dallas, TX, USA, ISBN: 0-7803-2934-1.
- [5] Broadwave technologies, "POWER DIVIDERS," *Available on internet*, 10.3.2012, http://www.broadwavetechnologies.com/CataloguesPDF/application_notes_power_dividers.pdf.
- [6] M. K. Amirhosseini, "Wideband differential phase shifter using microstrip nonuniform transmission lines," *Progress In Electromagnetics Research Letters*, Vol. 3, pp. 151 - 160, 2008.

- [7] RF, RFIC and Microwave Theory, Design, "Phase Shifter Design Tutorial," *www.rfic.co.uk*, *Available on internet*., 10.3.2012, http://www.odyseus.nildram.co.uk/RFMicrowave_Circuits_Files/Phase_%20Shifter.pdf.
- [8] Rogers corporation, "Rogers 4003C datasheet," *Available on internet*., 10.3.2012, <http://www.rogerscorp.com/documents/726/acm/RO4000-Laminates—Data-sheet.aspx>.
- [9] Mini-Circuits, "Bandpass Filter BFCN-7900+ datasheet," *Available on internet*., 10.3.2012, <http://www.minicircuits.com/pdfs/BFCN-7900+.pdf>.
- [10] Mini-Circuits, "Low Pass Filter LFCN-3400+ datasheet," *Available on internet*., 10.3.2012, <http://www.minicircuits.com/pdfs/LFCN-3400+.pdf>.

Domain Walls Contribution to the Magneto-Impedance Effect in Amorphous Magnetic Wire.

¹Viktória ŠUHAIJOVÁ (1st year)
Supervisor: ²Ján ZIMAN

¹Dept. of Physics, FEI TU of Košice, Slovak Republic

²Dept. of Physics, FEI TU of Košice, Slovak Republic

¹viktoria.suhajova@tuke.sk, ²jan.ziman@tuke.sk

Abstract — The magneto-impedance effect in the frequency range from 1 kHz to 1 MHz in amorphous ferromagnetic $\text{Co}_{68.2}\text{Fe}_{4.3}\text{Si}_{12.5}\text{B}_{15}$ wire with small negative magnetostriction was experimentally studied. The experiment presented gave possibility to measure contribution of domain walls to magneto-impedance in ferromagnetic wire with circular easy axis. It was shown that this contribution is relatively high (39.1 % at 1 MHz) and should be taken into account in theoretical interpretation or technical application of magneto-impedance in this type of materials.

Keywords — magneto-impedance, circular magnetization, domain wall.

I. INTRODUCTION

Giant magneto-impedance (GMI) effect is intensively studied during last two decades. The great sensitivity of the effect to very low values of external magnetic field makes it very promising for technological applications [1].

The magneto-impedance effect is defined as a change in the real and imaginary components of impedance due to external magnetic field. It is known that the AC current can flow only through the thin surface layer after achieving a sufficiently high frequency. This phenomenon is known as skin effect.

Skin depth δ is decreased with growing frequency of AC current f , circumferential magnetic permeability μ_ϕ and electrical conductivity of the material σ [2].

$$\delta = \sqrt{\frac{\rho}{\pi \cdot f \cdot \mu_\phi}}, \quad \rho = \frac{1}{\sigma} \quad (1)$$

II. EXPERIMENT

The measurements were carried out on as quenched amorphous ferromagnetic $\text{Co}_{68.2}\text{Fe}_{4.3}\text{Si}_{12.5}\text{B}_{15}$ wire with nominal diameter of 125 μm .

The experimental set-up used for the measurement of magneto-impedance presented in this paper is depicted in Fig.1. The ferromagnetic sample was in series with normal resistor R_N and the power supply of alternating voltage. The alternating voltage on sample (U_2 , channel 2 CH2) and also on

normal resistor (U_1 , channel 1 CH1) were measured by digital oscilloscope. Magneto-impedance Z was calculated using equations:

$$I_{\text{mean}} = \frac{U_{1\text{mean}} - U_{2\text{mean}}}{R_N}$$

$$Z = \frac{U_{2\text{mean}}}{I_{\text{mean}}} \quad (2)$$

where $U_{2\text{mean}}$ is the mean value of the voltage and I_{mean} is the mean value of the flowing current.

The experimental set-up used for the measurement of circular magnetization is depicted in Fig.2. The keys K_1 and K_0 made possible either the sample to be connected to power supply and to be magnetized circularly by magnetic field created by electric current flowing through it or the ends of the wire to be connected to the input of integrating amplifier (IA) and loops of circular magnetic flux versus axial magnetic field could be measured [3]. As is shown in Figs.1, 2, axial magnetic field was created by pair of Helmholtz coils.

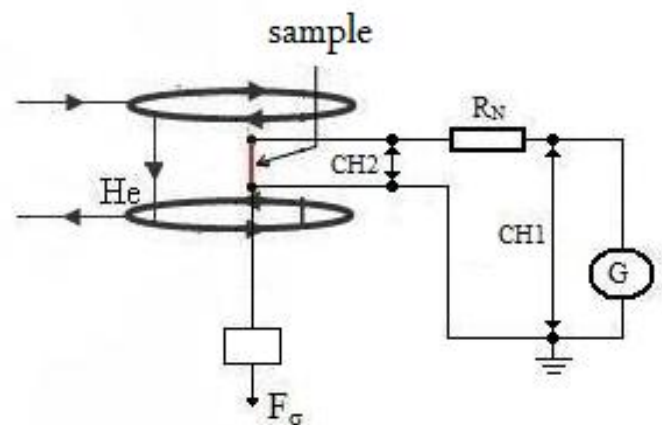


Fig.1. Experimental setup: He – Helmholtz coils, R_N – Normal resistor, G – power supply of alternating voltage, CH1, CH2 – channel 1, 2 of the digital oscilloscope, F_σ – stretching force.

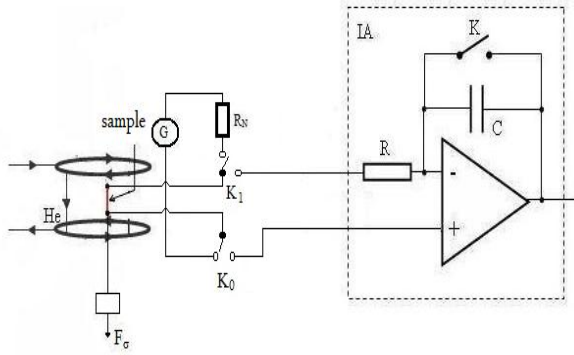


Fig.2. Experimental setup: He – Helmholtz coils, R_N – Normal resistor, G – power supply of alternating voltage, K_1 and K_2 – switching keys, IA – integrating amplifier, F_σ – stretching force.

The experimental set-up and experimental procedures proposed for magneto-impedance and circular magnetization measurements to make possible to measure both quantities under the same conditions.

The length of the sample used was 4.2 cm. It was placed in a sample holder which enabled torsion and tensile stresses to be applied simultaneously. The sample has low negative magnetostriction. Then the circular anisotropy could be increased by applied tensile stress. Applied torsion stress gave possibility to minimize dominant helical anisotropy which is present in the wire due to quenching process in rotating water.

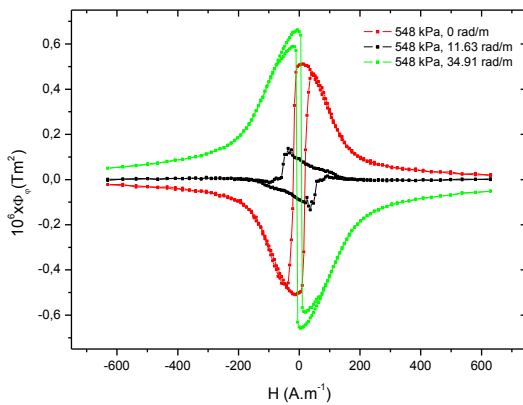


Fig.3 Axial hysteresis loop of the circular magnetic flux for different torsion.

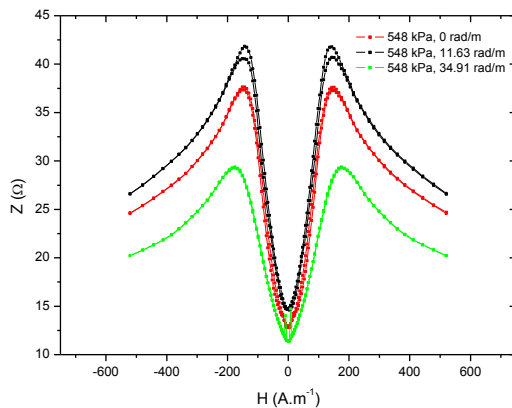


Fig.4. Hysteresis loops of magneto-impedance for different torsions.

III. RESULTS AND DISCUSSION

Hysteresis loops of magneto-impedance and circular magnetization as a function of axial field are shown in Figs.3, 4. They demonstrate a typical influence of torsion on both types of loops. We observe non-zero circular magnetization on the loop without torsion. It means that there is dominant helical anisotropy in the as-quenched sample. This anisotropy can be partly compensated by applied torsion. We could not achieve the situation when circular magnetization would be zero on the whole loop. In other words the abovementioned helical anisotropy could not be fully compensated by magnetoelastic anisotropy created by torsion. It was possible to achieve the state with minimum circular magnetization as it is demonstrated by the loop measured for torsion of 11.73 rad/m in Fig. 3. Further increase of torsion causes that opposite helical anisotropy is induced (see the loop for torsion of 34.91 rad/m in Fig. 3.).

Amplitude of magneto-impedance is the maximum for the case when helical anisotropy is minimum (torsion of 11.73 rad/m) as can be seen in Fig.4. Interesting is the fact that the highest magneto-impedance is observed for this case also for magneto-impedance in remanent state (zero axial field).

It is reasonable to expect that the number of circular domains in remanent state is increased by compensation of helical anisotropy. Experiments presented in this paper are to support this assumption and to measure domain wall contribution to the magneto-impedance at different frequencies.

Each point on dependences shown in Figs.5, 6 was measured using the following procedure. In the first step axial field created by Helmholtz coils changed from negative saturation to zero. Then a dc current I_{dc} was switched on and flew through the wire for about 1 second. This current created circular magnetic field in the wire. This field on the wire surface is given by formula

$$H_\phi = \frac{I}{\pi d} \quad (3)$$

where d is wire diameter.

After switching off circular field magneto-impedance or circular magnetic flux was measured.

Magneto-impedance as a function of H_ϕ measured at different frequencies and amplitudes I_m of ac current are shown in Fig.5.

We can see that at very low frequencies (1 kHz – 10 kHz), the magneto-impedance effect manifests itself mainly at higher values of amplitude of the AC current. Skin depth is greater than wire radius for these frequencies and changes of impedance are caused by a circular magnetization reversal.

More interesting magneto-impedance effects can be observed at higher frequencies (10 kHz – 1 MHz). The magneto-impedance versus circular field H_ϕ curves has typical shape for all frequencies. In the low field region a characteristic plateau can be observed. At circular field of about 2 A/m rapid increase of magneto-impedance starts and finally for the field higher than about 4 A/m magneto-impedance becomes constant. Very similar behavior can be observed on circular magnetization flux versus the field H_ϕ curve shown in Fig.6. The same regions can be identified also on this curve. Changes of magneto-impedance are accompanied by changes of circular magnetization. Observed behavior can be interpreted in the following way. Reduction of helical anisotropy results in a relatively high number of

circular domains and so also in a great number of domain walls in remanent state after switching off axial field. In low circular field region reversible movement of domain wall causes that no changes in circular magnetization is observed. Since there is no change in magnetic state no changes are observed also on magneto-impedance curves. For circular field higher than about 2 mA circular domain wall displacements become irreversible and some of them even annihilate. In other words the number of domain walls is reduced in this region. This reduction of domain walls is accompanied by the decrease of magneto-impedance. For circular field higher than about 4 A/m all walls annihilated and no changes are observed on both magneto-impedance and circular magnetization curves.

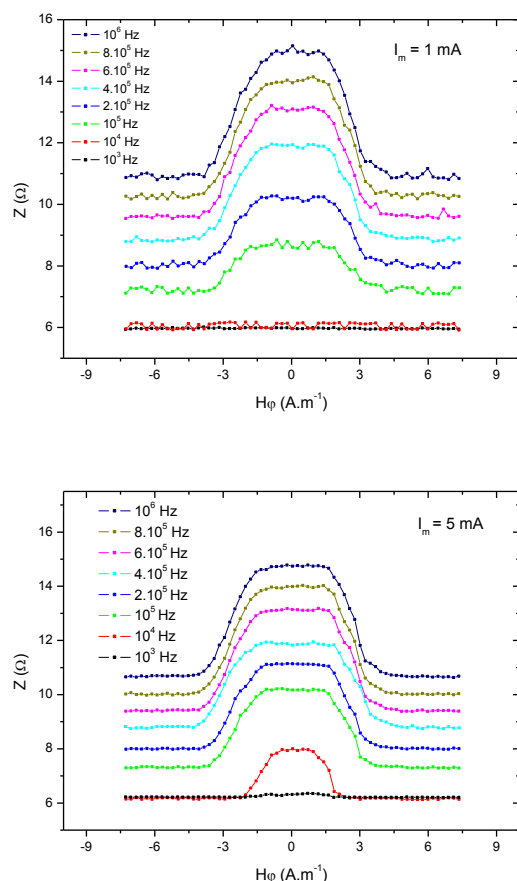


Fig.5. Impedance Z versus circular magnetic field H_ϕ for amplitudes of AC current 1 mA and 5 mA.

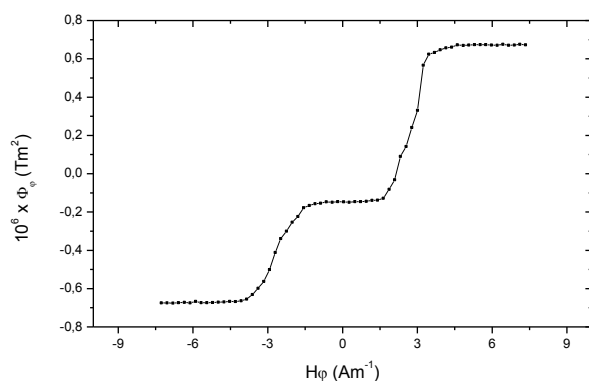


Fig.6. Circular magnetization flux Φ_ϕ as a function of circular field H_ϕ .

IV. CONCLUSION

We can conclude that the experiment presented in this paper gave possibility to measure contribution of domain walls to magneto-impedance in ferromagnetic wire with circular easy axis. It was shown that this contribution is relatively high (39.1 % at 1MHz) and should be taken into account in theoretical interpretation or technical applications of magneto-impedance in this type of materials.

V. ACKNOWLEDGMENT

The research presented in this article was supported by S.G.A Project No. 1/0778/12.

The author would like to acknowledge her supervisor doc. RNDr. Ján Zíman CSc. for proposing the topic of this article and for continuous help.

REFERENCES

- [1] M. Knobel, M. Vázquez, L. Kraus, *Handbook of Magnetic Materials* (Book style). Elsevier Science B.V.: 2003, pp. 497-559.
- [2] D. Jiles, *Introduction to Magnetism and Magnetic Materials* (Book style). 2nd ed., Suffolk: 1998, pp. 57-60.
- [3] J.Zíman, B.Zagyi, "DC-magnetoresistance in surface crystallized FeSiB amorphous wire" *J. Magn. Mater.* 169(1997) pp. 98-104

Effect of electromagnetic field

¹Marek PAVLÍK (1st year), ²Ján ZBOJOVSKÝ (1st year)
Supervisor: ³Iraida KOLCUNOVÁ

^{1,2,3}Dept. of Electric Power Engineering, FEI TU of Ko-ice, Slovak Republic

¹marek.pavlik@tuke.sk, ²jan.zbojovsky@tuke.sk, ³iraida.kolcunova@tuke.sk

Abstract With continuously increasing number of devices working on principles of electromagnetic field, more emphasis is put on biological effects electromagnetic field has. In present time effects of electromagnetic field on living organisms are treated by World Health Organization (WHO), International Commission on Non-Ionizing Radiation Protection (ICNIRP). They stated that electric devices must follow certain limits, they have to be electromagnetically compatible in order to be used in practice without being threat to human health

Keywords electromagnetic field, mobile phone, source of electromagnetic field.

I. INTRODUCTION

In Slovak republic this issue is addressed in Act 534/2007 Collection ó Edict of Ministry of Health of the Slovak Republic from 16 August 2007 about details and requirements on sources of electromagnetic radiation and on limits of exposure to electromagnetic radiation of population in environment. Edict of Ministry of Health of the Slovak Republic defines exposure action values. Allowable exposure action values are presented in Table I. while f stands for frequency as stated in transmission bandwidth column.

TABLE I
Allowable exposure action values[2]

Frequency range	electric field intensity E (V/m)	magnetic field intensity H (A/m)	magnetic induction B (μT)
0 Hz to < 1 Hz	-	$3,2 \cdot 10^{-4}$	$4 \cdot 10^{-4}$
1 Hz to < 8 Hz	10 000	$3,2 \cdot 10^4/f^2$	$3,2 \cdot 10^4/f^2$
8 Hz to < 25 Hz	10 000	$4\,000/f$	$5\,000/f$
0,025 kHz to < 0,8 kHz	$250/f$	$4/f$	$5/f$
0,8 kHz to < 3 kHz	$250/f$	5	6,25
3 kHz to < 150 kHz	87	5	6,25
0,15 MHz to < 1 MHz	87	$0,73/f$	$0,92/f$
1 MHz to < 10 MHz	$87/f^{1/2}$	$0,73/f$	$0,92/f$
10 MHz to < 400 MHz	28	0,073	0,092
400 MHz to < 2 000 MHz	$1,375 \cdot f^{1/2}$	$0,0037 \cdot f^{1/2}$	$0,0046 \cdot f^{1/2}$
2 GHz to 300 GHz	61	0,16	0,20

Directly measurable quantities that characterize the electromagnetic field components are:

- electric field intensity E [V/m] ó component of electric field

- magnetic induction B [T] ó component of the magnetic field
- magnetic field intensity H [A/m] ó component of the magnetic field

Derived quantities characterizing the effects of electromagnetic fields on human health are:

- power flux density S [W/m²]
- current density J [A/m²]
- Induced current I_L [mA]
- Specific absorption rate SAR [J/kg.s].

Based on the above Edict can be made so. special values of continuous exposure to electromagnetic fields generated by the emission. The population of Slovakia are the following variables and their special values:

1. Electric field strength E , which are action values of E_A for the frequency from 400 MHz to <2000MHz determined from the relation (1)

$$E_A = 1,375 \sqrt{f_{MHz}} \text{ [V/m]} \quad (1)$$

where: f_{MHz} is frequency of radiated electromagnetic energy [MHz].

2. Magnetic field intensity H , which are action values of H_A for the frequency from 400 MHz to <2000MHz determined from the relation (2)

$$H_A = 0,0037 \sqrt{f_{MHz}} \text{ [A/m]} \quad (2)$$

3. Magnetic induction B , which are action values of B_A for the frequency from 400 MHz to <2000MHz determined from the relation (3)

$$B_A = 0,0046 \sqrt{f_{MHz}} \text{ [T]} \quad (3)$$

II. BIOLOGICAL EFFECTS OF MAGNETIC FIELD

In the world there are plentiful of statistics and experiments that show link between electromagnetic field and living organisms. Biological effects of electromagnetic field can be sorted into

- Thermal effects that manifest themselves by rise of tissue temperature, in extreme cases can cause flash burns
- Non-thermal effects, after long term exposure can cause feeling of weakness, sleeping disorders, memory disorders.

III. SOURCES OF ELECTROMAGNETIC RADIATION

Currently the most wide-spread sources of electromagnetic radiation are mobile telephones (MT) and wifi devices. Seeing that present time brought us a luxury in form of using wireless internet, sources of electromagnetic radiation increased in numbers.

IV. EFFECTS OF MOBILE TELEPHONES AND THEIR ELECTROMAGNETIC RADIATION ON HUMAN HEALTH

Mobile telephones are wireless receivers and transmitters. According to recommendation of European Union they have to be designed in a way so they will not exceed limit values of human exposure to high-frequency radiation. Doc. Ing. Zdeněk Matoušek, PhD., Ing. Mikuláš Trštronek, PhD. and Doc. Ing. Jozef Jakub, PhD. of Armed Forces Academy of Generál Milan Rastislav Štefánik - department of electrotechnics describe effects of electromagnetic radiation of mobile telephones on human organism in their article.[2]

Simple measurement was conducted, it was consisting of placing measured mobile telephone (MT) in the range of 3 meters from measuring antenna. Measuring was conducted in attenuation chamber, so the particular MT would emit maximum capacity signal. Ten random MT samples by various manufacturers were used for measuring. These measurement gave values showed in Fig. 1 for 900MHz bandwidth and in Fig. 2 for 1800MHz bandwidth. [2]

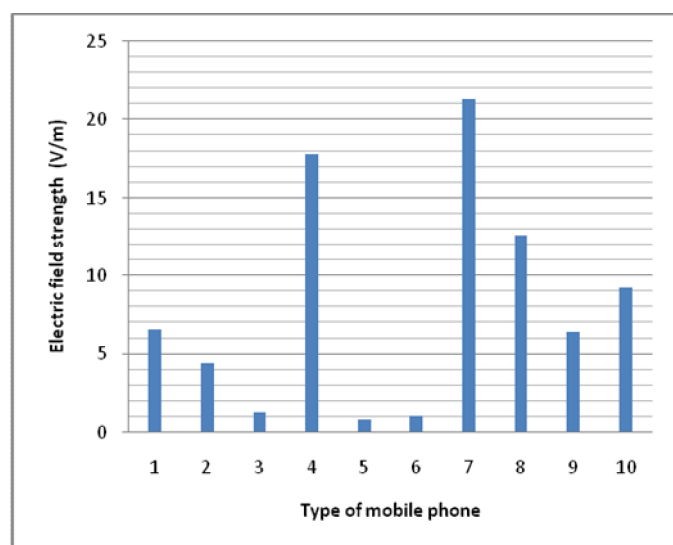


Fig. 1 Electric field strength 10 randomly selected mobile phones to 900MHz[2]

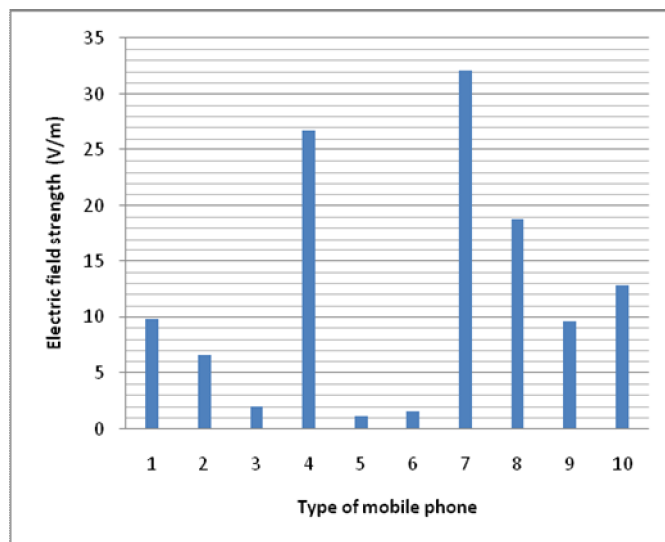


Fig. 2 Electric field strength 10 randomly selected mobile phones to 1800MHz[2]

From values stated higher follows that all MT correspond with the limit values of intensity of electric field defined in the Edict of Ministry of Health of the Slovak Republic. However considering these measurements, it is needed to state that almost with every tested MT's allowed exposition of electromagnetic field value was exceeded after 50 minutes of voicecall. In two cases allowed exposition of electromagnetic field value was exceeded after 11 or 14 minutes.

In time of frequent usage of MT it is not so rare to have a voicecall lasting more than 30 minutes, especially in some sectors. Survey was conducted on a sample of 100 people. They were asked single question ó In average, how long does your voicecall last. Result of this statistic is shown on Fig. 3 as relation of number of people to lasting of voicecall.

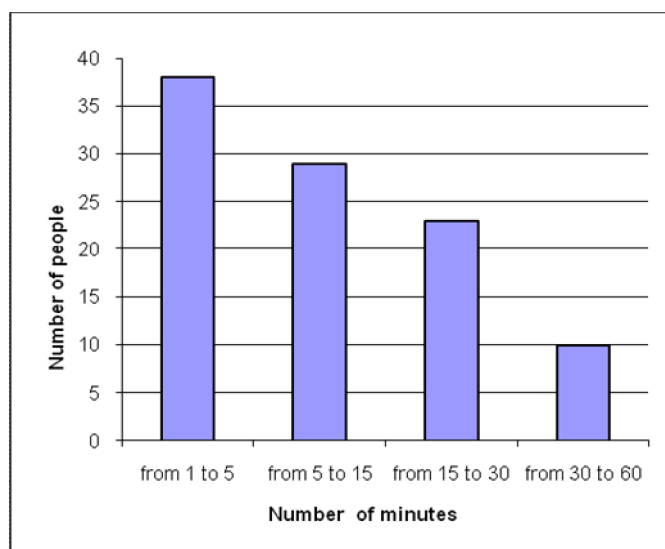


Fig. 3 The number of people depending on the length of their common call made

Ten people out of one hundred have voicecalls lasting from 30 to 60 minutes. If we speculate that all of asked use one of two MT that were proved to exceed allowed exposure of magnetic field value after 11 ó 14 minutes, it would mean that approximately 60% of population is exposed to excessive electromagnetic field values. It is needed to take into

consideration that measurements were conducted on older types of MT. We can presume that with newer types of MT the exceeding of limits shouldn't be so significant. Anyway it is important to remind that mobile telephones have effect on human organism even if it is often discussed topic in the wide scientific community.

V. EFFECTS OF WI-FI DEVICES ON HUMAN HEALTH

In the TV programme Panorama that aired 21 May 2007 on British BBC, reporters drew attention to exceeded radiation values of wifi connection at one of the British schools. Measured values reportedly exceeded the values of mobile telephones three times. Wi-Fi connection to internet is mainly used by laptop users and is common in whole Europe. Chairman of Health Protection Agency (HPA) William Stewart called for reevaluating of wi-fi connection to internet during debate about possible malignance of airwaves to human health. According to Stewart devices such a mobile telephones and wi-fi have negative effects on human health. Scientific community however strongly disagrees. "It is improbable that wi-fi would be harmful to human health" said professor Lawrie Challis from Nottingham University. According to Guardian, scientists labeled Panorama as "roughly unscientific" and "intimidating". "Wi-Fi is a technology that uses airwaves with very low intensity. Their wave length is similar to one used in homes in microwaves, but have hundredthousand times lower intensity," explained for BBC specialist for medical physics ó professor Malcolm Sperrin. Radiation emitted by airwaves (wi-fi including), visible radiation, microwaves and mobile telephones can at high dosage warm up human tissue and so damaging it. There is no proof of this happening at low dosage. [3]

Airwaves and other non-ionizing radiations are part of our lives for more than hundred years and yet there has been no record of its negative effect on human health" Sperrin added. [3]

Alle Johanson from Swedish Karolinska institute, who appeared in aired TV programme, thinks otherwise. According to her, it was recorded many times before that even weak radiation can have effect on human health and can for example damage chromosomes bearing genetic informations. Similar effect was allegedly observed by Henry Lai from Washington State University. Scientists generally claim that waves used by wi-fi for data transfer are safer than GSM connection of mobile phones due to working on shorter distances and not requiring so much energy. According to HPA after a year of constant connection to internet through wi-fi human body absorbs as much radiation as after 20 minutes of MT voicecall. [3]

It is questionable if the values would not be exceeded if the effect of more devices simultaneously in action was measured. [3]

VI. THERMAL EFFECTS OF ELECTROMAGNETIC FIELD

Biological effects that result from heating of tissue by RF energy are often referred to as "thermal" effects. It has been known for many years that exposure to high levels of RF

radiation can be harmful due to the ability of RF energy to heat biological tissue rapidly. This is the principle by which microwave ovens cook food, and exposure to very high RF power densities, i.e., on the order of 100 mW/cm² or more, can clearly result in heating of biological tissue and an increase in body temperature. Tissue damage in humans could occur during exposure to high RF levels because of the body's inability to cope with or dissipate the excessive heat that could be generated. Under certain conditions, exposure to RF energy at power density levels of 1-10 mW/cm² and above can result in measurable heating of biological tissue (but not necessarily tissue damage). The extent of this heating would depend on several factors including radiation frequency; size, shape, and orientation of the exposed object; duration of exposure; environmental conditions; and efficiency of heat dissipation. [4]

Studies have shown that environmental levels of RF energy routinely encountered by the general public are far below levels necessary to produce significant heating and increased body temperature. However, there may be situations, particularly workplace environments near high-powered RF sources, where recommended limits for safe exposure of human beings to RF energy could be exceeded. In such cases, restrictive measures or actions may be necessary to ensure the safe use of RF energy. [4]

In addition to intensity, the frequency of an RF electromagnetic wave can be important in determining how much energy is absorbed and, therefore, the potential for harm. The quantity used to characterize this absorption is called the "specific absorption rate" or "SAR," and it is usually expressed in units of watts per kilogram (W/kg) or milliwatts per gram (mW/g). In the far-field of a source of RF energy (e.g., several wavelengths distance from the source) whole-body absorption of RF energy by a standing human adult has been shown to occur at a maximum rate when the frequency of the RF radiation is between about 80 and 100 MHz, depending on the size, shape and height of the individual. In other words, the SAR is at a maximum under these conditions. Because of this "resonance" phenomenon, RF safety standards have taken account of the frequency dependence of whole-body human absorption, and the most restrictive limits on exposure are found in this frequency range (the very high frequency or "VHF" frequency range). [4]

Although not commonly observed, a microwave "hearing" effect has been shown to occur under certain very specific conditions of frequency, signal modulation, and intensity where animals and humans may perceive an RF signal as a buzzing or clicking sound. Although a number of theories have been advanced to explain this effect, the most widely-accepted hypothesis is that the microwave signal produces thermoelastic pressure within the head that is perceived as sound by the auditory apparatus within the ear. This effect is not recognized as a health hazard, and the conditions under which it might occur would rarely be encountered by members of the public. Therefore, this phenomenon should be of little concern to the general population. Furthermore, there is no evidence that it could be caused by telecommunications applications such as wireless or broadcast transmissions. [4]

It has been claimed that some parts of the human head are more sensitive to damage due to increases in temperature, particularly in anatomical structures with poor vasculature, such as nerve fibers. More recent results from a Swedish scientific team at the Karolinska Institute (Lonn, Ahlbom, Hall and Feychting) have suggested that continuous use of a mobile phone for a decade or longer can lead to a small increase in the probability of getting acoustic neuroma, a type of brain tumor. The increase was not noted in those who used phones for less than 10 years. The study has been criticized for possible problems in data analysis such as recall bias. However, another study conducted by the Swedish National Institute for Working Life supported an increased risk of malignant tumors on the side of the head the phone is used. Such long term heavy use involved phones of older higher power analog designs; that were first introduced to Sweden in 1984, earlier than many other countries. [4]

VII. MEASURING THE ELECTROMAGNETIC FIELD

The human body is constantly exposed to the impact of electromagnetic fields. However, it is not always exposed to direct impact. Since most of the time a person spends inside (in the office, at home), is exposed to electromagnetic fields, which are damped as walls of buildings. It is important to know the throughput of electromagnetic waves of different materials. To determine the need to take measurements. The measurement consists of a measured resource of the electromagnetic field of the wall, respectively. barriers and the spectrum analyzer. The whole measurement should be done in a Faraday cage that was prevented by the influence of other sources of electromagnetic fields. Configuration of the workplace the measurement of electromagnetic field is in Fig. 4.

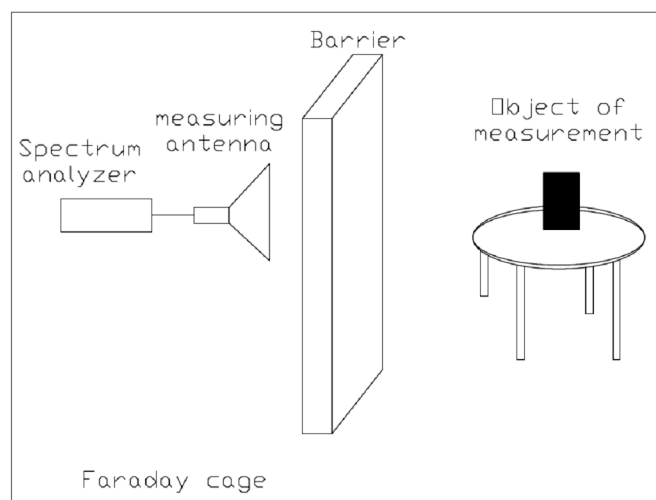


Fig. 4 Configuration of the workplace the measurement of electromagnetic field

VIII. CONCLUSION

Effect of electromagnetic field on human organism is a very hot topic and there has not been clearly established whether the electromagnetic fields affect the human body or not. In any case, it is certain that we are daily exposed to the influence and most likely will continue. It is therefore necessary to examine the effect of a measure.

ACKNOWLEDGMENT

This publication is the result of the Project implementation: Protection of the population of the Slovak Republic from the effects of electromagnetic fields, ITMS: 26220220145, activity 2.1 supported by the Research & Development Operational Programme funded by the ERDF .



We support research activities in Slovakia / Project is cofinanced from EU funds.

REFERENCES

- [1] *Edict of Ministry of Health of the Slovak Republic from 16.august 2007:* for details on requirements for sources of electromagnetic radiation and exposure limits for electromagnetic radiation of the population in the environment.
- [2] Z. Matoušek, M. Trštronek, J. Jakub , *Effect of electromagnetic mobile phone radiation on human body*, 2009, pp. 73680: Available on the internet: <http://www.aos.sk/spe/seminare/SPE_2009/zbornik/07matousek_09.pdf> .
- [3] *BBC haunted harmfulness of mobile devices* in Journal Pravda [online]. 23/5 2007. Available on the internet: <http://spravy.pravda.sk/bbc-strasi-skodlivostou-mobilnych-zariadeni-fnp-sk_zaujima.asp?c=A070523_083249_sk_zaujima_p23>.
- [4] R. F. Cleveland, Jr., J. L. Ulcek *Questions and answers about biological effects and potential hazard of Radiofrequency electromagnetic fields*, August 1999, fourth edition.

Effective Measurement Stand for Teaching Basic Electronic Circuits

¹Marek GODLA(2st year), ²Daniel FABRI(2nd year), ³Jozef LIPTÁK(1st year)
Supervisors: ¹Linus MICHAELI, ^{2,3}Ján ŠALIGA

^{1,2,3,4}Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

¹marek.godla@tuke.sk, ²dfabri@azet.sk @tuke.sk, ³jozef.liptak @tuke.sk

Abstract— the paper presents project and outcomes from remote access laboratory for experimental exercises. This project has been implemented in the course „Basics of electronics“. The experimental laboratory stands are controlled remotely from client's computers across the Internet. User can change configuration of circuit components, component values and also select measuring nodes. The control and data presentation is realized using virtual instruments implemented in dynamic web pages broadcasted by laboratory web server with installed plug-in “Run Time Engine” by National Instruments. Before performing experimental measurement student have to proceed in the three steps. In the first step students have to be familiar with circuit theory and simulation of measured circuit. The second step is real measure on this circuit. Finally in the last step students have to compare result from simulation and from the real measurement. This project has been created in programming environment LabVIEW. We have used for minimize realization cost DAQ cards instead of traditional stand-alone instrumentations.

Keywords - remote laboratory; LabVIEW, virtual instrumentation, education, electronic circuits

I. INTRODUCTION

The main task of experimental laboratory stands in the courses of electronic is obtaining practical and theoretical knowledge. Practical experiments allow students familiarizing with measuring instruments, assessment of accuracy of simulation results in comparison with measured ones and enhance the deepness and quality of achieved knowledge in the subject. For this solution it's possible to realize real measurement and verification of the acquired theoretical knowledge in laboratory. This approach requires the personal presence of a supervisor in the laboratory especially for the students starting the bachelor level of education and also restricts the laboratory access time according to university internal security and other conditions. The basis way when it's possible to get are two solution of this problem. One is virtual laboratory that is based on simulation . The main disadvantage of this approach is inaccuracy of the behaviour model of real measurement. The other way is used solution from remote laboratory based on real hardware. Design of remote laboratories for various application fields was studied by many authors, e.g., [1] – [3].

The aim of this paper is to describe hardware and software solution for teaching subject “Basic of Electronics” using real measuring workplace accessible across the Internet developed at the Department of Electronics and Multimedia Communications, Technical University of Kosice.

II. IMPLEMENTATION OF EXPERIMENTAL WORKS IN CURRICULA

The simple way in effective teaching is focused on electronic circuits at bachelor and master level of university education is not possible without simulation and real laboratory experiment.

It is very different between simulation experiment and real experiment in real laboratory measurement. Electronic circuit simulation is very simple way how it's possible to verify theoretic knowledge without risk to damage electronic components and event measuring instrumentation. Main knowledge is use simulation in big groups of students. It enables better understanding of relations among electronic components, their values, connections and the final properties of resulting electronic circuits.

On the other hand, the simulations sometimes leads students to the expectation that behaviour of real circuit is always in perfect equality to the results of simple simulation and that the realized circuits will work exactly as it worked in simulation environment. This fact is the main reason why a real experimental measurement and testing of designed circuits seems to be the next necessary stage in teaching circuit synthesis.

Real experimental laboratory requires corresponding equipments for given experiments and adequate time schedule to enable students performing the measurements. In praxis the main limitation can be especially convenient time schedule, because according security reasons the presence of students in laboratory require also presents a tutor. On the other hand more advantage is for students who prefer to work late in the evening or during weekend when university laboratories are usually closed.

The Project of measure stand wants to solve mentioned problems and attempt to implement three stage chain “theory – simulation – experiment” solving all here above mentioned problems and limitations to teaching subject “Basics of Electronics” at the Department of Electronics and Multimedia communications, Technical university of Kosice.

The practical realization of stages is:

1. Stage “Theory”: students passing the theoretical lectures design basic circuits within numerical exercises.
2. Stage “Simulation”: The designed “calculated” circuits are simulated using PSpice tools.
3. Stage “Experiment”: The designed and simulated circuits are realized and measured. To save time and costs this stage is realized in remote access laboratory that is “open” 24 hour in a day and 7 days in a week.

III. HARDWARE ARCHITECTURE OF THE REMOTE ACCESS LABORATORY

The measuring stand contents real measuring instruments for simulate of real measure. Those instruments are being controlled locally by a stand/laboratory server. Instruments control commands and acquired data are collected/presented to user through dynamic web pages broadcasted by laboratory web server. A utilized web camera allows increasing the impression of real laboratory for user displaying image from front panels of instruments and view of whole measurement workplace. The instrumentation must be switch on 24 hour in a day for months what can lead especially to a damage of these instruments.

The alternative solution is utilization of multifunction data acquisition board (DAQ) installed directly in laboratory server. The board contain multiplexed analogue input channel with 16 inputs, that are used for acquiring analogue signal from circuit under test, 2 analogue outputs that can substitute two generators 2 -3 programmable counters that can be utilized as pulse generators or frequency meters and 8 – 16 programmable digital inputs/outputs that can control circuit under test by, e.g., controlling multiplexer changing component in the circuit, or they can be used for generation digital stimulus or acquiring digital signals from circuit under test.

Instrumentations based on DAQ cards require also software realizing data acquisition and presentation in the form of “virtual instrumentation”. Virtual instruments are very simply integrated to dynamic web pages. Also web server administrating the web pages can be simply implemented in the same machine like performs the laboratory – measurement server role. In other words one PC can perform control of a few measurements stands using DAQ cards and virtual instruments and serves in parallel as web server. Required realization costs could be decreased at least 5 – 10 times in comparison with traditional solution utilizing stand-alone instruments.

Laboratory hardware consists of two DAQ boards (PCI 60336 and PCI 6251) with connector block SCC-68 where circuits under test are connected. In the present there are five such circuits:

1. Stabilized voltage source
2. Photovoltaic cell
3. Differential and Darlington transistor amplifiers
4. Operational amplifiers
5. Power amplifier

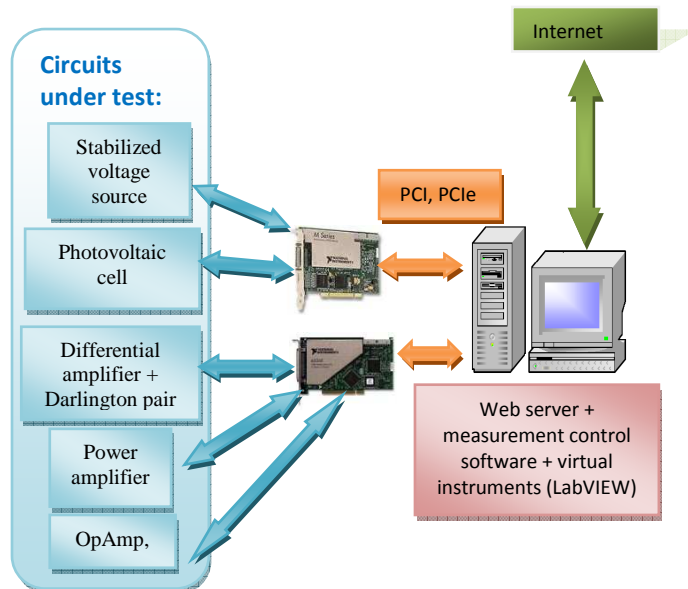


Figure 1. Architecture of the proposed remote laboratory

The present variety of circuits under test will be spread in near future by some new other circuits. The control PC is a common PC with Intel® Core™2 Quad processor, 4GB of memory and OS Windows XP. The PC performs both tasks – laboratory server as well as web server. The practical experiments showed that this simple solution powered by one simple low cost PC has high efficiency. When one student has control on the measuring stand processor works 10% of overall performance.

To increase the number of possible changes in component network of measured circuit, analogue multiplexers were implemented into the circuit. For example, the laboratory stand for testing operational amplifiers circuits consists of four independent circuits for four laboratory exercises. Each measuring circuits is being controlled by a multiplexer, which allow changing circuit components according the command by a student from remote computer. Control multiplexer as it is shown in Fig. 2 for inverting amplifier. The final value of feedback resistor is determined by parallel connection combination of four resistors connected through the analogue switching circuit.

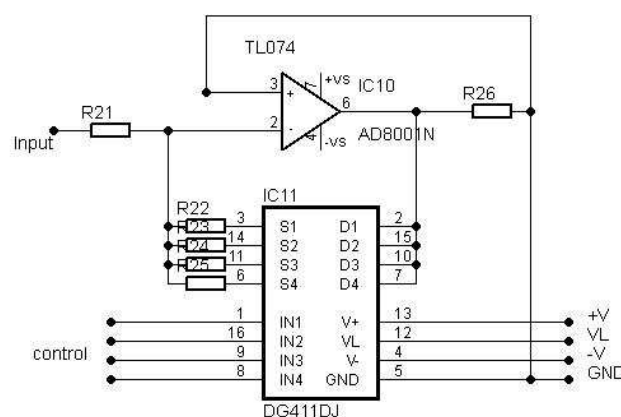


Figure 2. Switching of feedback resistor in inverting circuits of operational amplifier

IV. CONTROL SOFTWARE FOR MEASUREMENT STANDS

The control software has been developed in LabVIEW by National instruments. LabVIEW was chosen because of its features: easy learning and effective programming, rich libraries supporting DAQ cards, signal processing and

visualisation, etc. An additional advantage of LabVIEW is simple way of publish control program on the Internet. The only disadvantage of LabVIEW web server is that clients must have installed a special plug-in “Run Time Engine” that can be downloaded and installed to client web browser free [4].

LabVIEW web server also administrates the connection of every client. The first connected student asking for access to a measurement workplace gives the access automatically for unlimited time. Because we have five independent workplaces five students can work in parallel and independently. If another student connects to the server and asks for access to an occupied workplace, web server inserts this request to the queue of requests and limits the rest time for active user to a limit that could be set by administrator – usually 10-15 minutes. The active user is noticed that next user wants to take control of the measure stand and also is noticed that the workplace will be released in a time.

The design of front panels of VIs controlling particular workplaces was created in the form very similar to real instrumentation. User can control the virtual instruments as real ones and their behaviour is very similar to real traditional instruments.

A. Virtual oscilloscope

Program allows controlling basic settings:

- time base
- voltage range
- ON/OFF inputs
- triggering level
- V-position
- H- position
- coupling of channels
- button AUTO
- button RUN/STOP.

User interface of application software for control laboratory oscilloscope you can see in Fig. 3

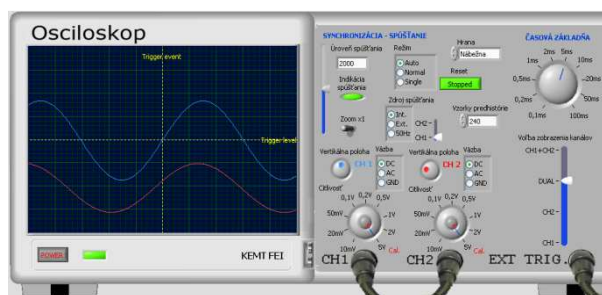


Figure 3. User interface of virtual oscilloscope

Software allows acquiring measured signals, shows measuring waves, enables analyzing of measuring waves (amplitude, period, minimal voltage, maximal voltage, frequency)

B. Virtual generator

User interface of virtual signal generator is shown in Fig. 4



Figure 4. User interface of virtual signal generator

- Software allows changing these setting:
- output waveform
- amplitude
- offset
- frequency
- additional noise

C. Virtual digital multimeter

User interface of virtual digital multimeter is shown in Fig. 5. This program allows controlling basic settings:

- selection of the measured quantity
- selection of measuring range
- selection of measuring mode



Figure 5. User interface of virtual digital multimeter

- D. Control of other signals, selection of measuring nodes, presenting some other measured values

In workplace the changing component values, selection of measuring nodes and control and displaying some other non electrical values. All of this is realised by user interface in the form of special virtual instruments. The typical example is user interface developed for workplace with photovoltaic cell (Fig. 6.). The user controls here photovoltaic cell lighting, changes load resistance and measures light intensity by luxmeter realised by photoresistor.

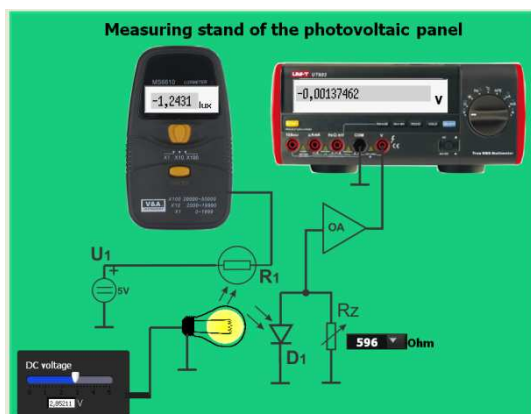


Figure 6. User interface of photovoltaic cell workplace

Another example is in Fig. 7. User can here set up the load of stabilised voltage source on the workplace.

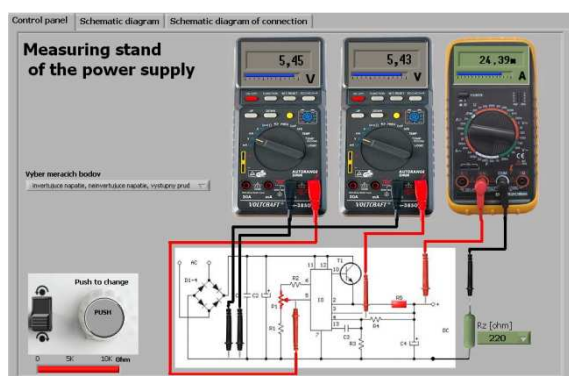


Figure 7. User interface of power supply measure stand

V. CONCLUSIONS

Realization of remote controlled measurement laboratory for measuring basic electronic circuits as a support and the last stage of chain “theory – simulation – experiment” for effective teaching subject “Basics of Electronics” at the Department of Electronics and Multimedia communications, Technical university of Kosice is the process of testing, improvement and spreading in presence. In the summer semester of 2011 it is applied to regular teaching process at the department for the first time.

In the future measure workplace will be enriched by testing stand of ADC converters, where will real AD converter connected to FPGA circuit which save the testing time and help users better understand of problematic of the testing ADC.

ACKNOWLEDGMENT

The work is a part of project supported by the Educational Grant Agency of Slovak republic (No. 3/7115/09)

This work was supported by Agency of the Ministry of Education of the Slovak Republic for the Structural Funds of the EU under the project Development of Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030).

REFERENCES

- [1] Daponte, P, Grimaldi, D., Rapuano S.: New Frontier of the m-Learning in Instrumentation and Measurement: Experiment Creation, Proc. Of 16th International Conference on Software, Telecommunications and Computer Networks, Split, CROATIA, SEP 25-27, 2008, pp. 412-416
- [2] L. Peretto, S. Rapuano, M. Riccioc, D. Bonattia: Distance learning of electronic measurements by means of measurement set-up models, Measurement, Volume 41, Issue 3, April 2008, Pages 274-283
- [3] D. Sampera, J. Santolaria, A. Majarena and J. Aguilar: Comprehensive simulation software for teaching camera calibration by a constructivist methodology, Measurement Volume 43, Issue 5, June 2010, Pages 618-630
- [4] <http://joule.ni.com/nidu/cds/view/p/id/2088/lang/en>

Electromagnetic processes of switched reluctance motor with variable phase inductance

¹Nataliia ISTOMINA, ²Matúš OCILKA (1st year), ³Ján MOLNÁR

¹Institute of Electromechanics, Energy Saving and Control Systems,
Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine

^{2,3}Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

¹istominaNM@yandex.ua, ²matus.ocilka@tuke.sk, ³jan.molnar@tuke.sk

Abstract— The mathematical model of switched reluctance machine with variable phase inductance was received. This model has allowed to analyze the variable inductance influence on the phase current transient process. It permits to increase the mathematical model accordance to physical processes in the switched reluctance machine.

Keywords— switched reluctance machine, variable phase inductance, rotor rotation angle.

I. INTRODUCTION

Recently, switched reluctance drive research were activated due to intensive development of power and control electronics [1], [2].

Analytic representation of switched reluctance motor (SRM) complicated by significant nonlinearities [2]. The SRM mathematical model should consider variable phase inductance. In the known works [3], [4] inductance described as function of rotor angle:

$$L(\varphi) = \begin{cases} L_{\min}, & \varphi < 0; \\ L_{\min} + K\varphi, & 0 < \varphi < \beta_s; \\ L_{\max}, & \beta_s < \varphi < \beta_r; \\ L_{\max} - K(\varphi - \beta_s + \beta_r), & \beta_s < \varphi < \beta_s + \beta_r; \\ L_{\min}, & \varphi > \beta_s + \beta_r \end{cases} \quad (1)$$

where L_{\min} – inductance minimum value, H;

L_{\max} – inductance maximum value, H;

β_s – stator pole pitch, degree;

β_r – rotor pole pitch, degree;

K – slope coefficient;

φ – rotor angle, degree.

The function represented on fig. 1.

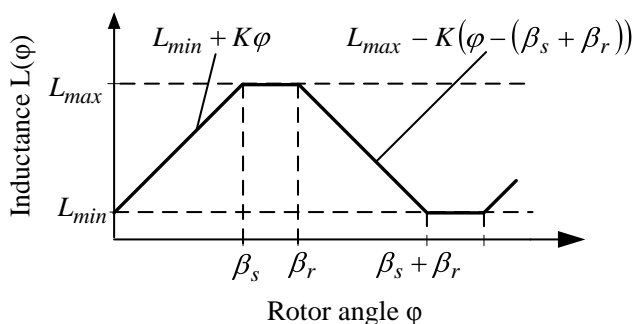


Fig. 1. Inductance dependence on rotor angle.

This inductance representation is not considered speed of angle rotor changing. Notably slope coefficient is constant and determined by the formula:

$$K = \frac{L_{\max} - L_{\min}}{\beta_s} \quad (2)$$

The actual problem is the motor inductance description with time dependence and rotor angle derivative take into account.

II. OBJECTIVE

Improving the accuracy of mathematical modeling of switched reluctance motor.

III. MATERIALS AND FINDINGS

Inductance form determined by the stator-rotor pole pitch ratio.

Where are two types of this ratio:

–case 1 – $\beta_s = \beta_r$;

–case 2 – $\beta_s < \beta_r$.

Consider the stator-rotor pole pitch ratio influence ratio on inductance form, on the SRM 6/4 example. Below the simulated cross-section presented: for case 1 – fig. 2; for case 2 – fig. 3.

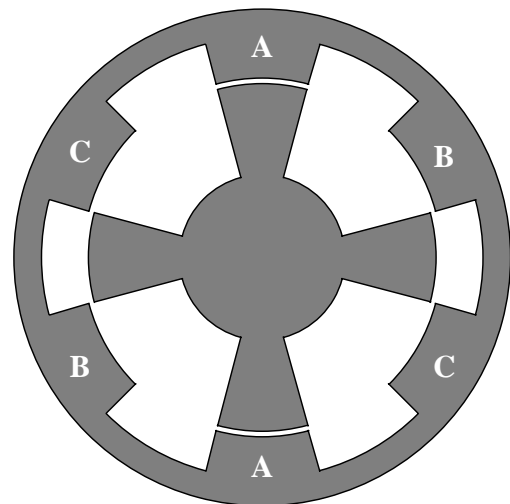
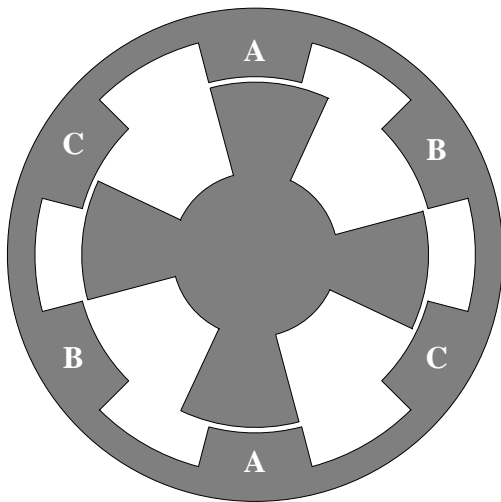
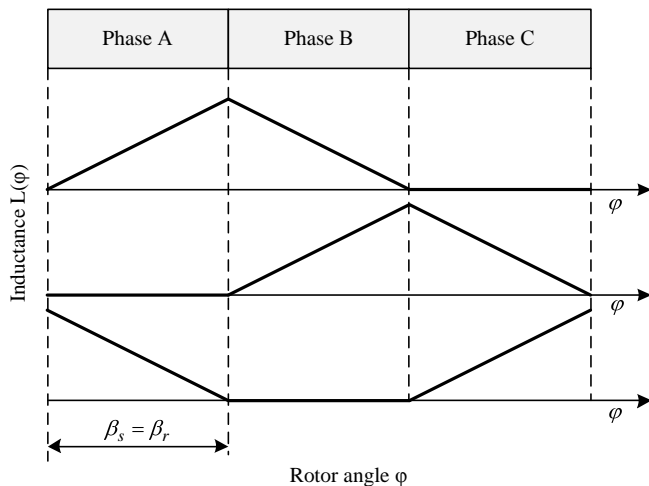


Fig. 2. Simulated cross-section for $\beta_s = \beta_r$.

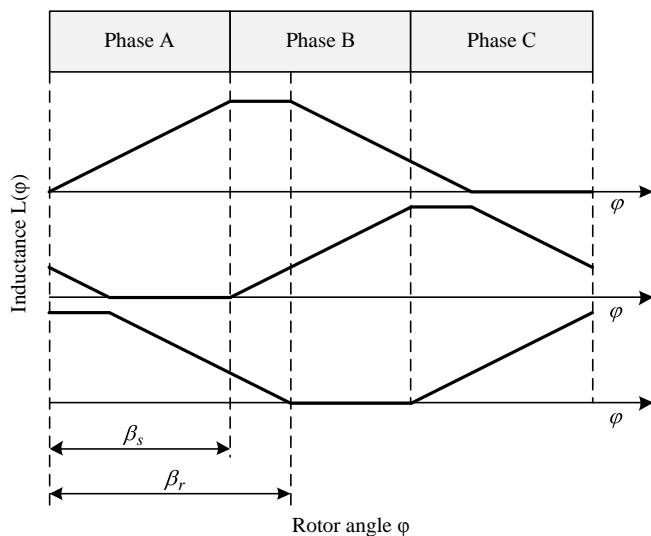
Fig. 3. Simulated cross-section for $\beta_s < \beta_r$.

The phases switched on the moment fit to inductance maximum value. This corresponds to the maximum overlapped area of stator and rotor poles.

In *case 1* phase inductance has triangle form (Fig. 4).

Fig. 4. Phase inductance form for $\beta_s = \beta_r$.

In *case 2* phase inductance has form as shown on fig. 5.

Fig. 5. Phase inductance form for $\beta_s < \beta_r$.

The case 2 inductance form provides torque increasing time and vibrations reducing. Therefore, further researches were conducted for this case.

We proceed to the time dependent of rotor angle $\varphi(t)$.

It is known that the angular velocity of rotated object and the angle related by the expression:

$$\omega = \frac{d\varphi}{dt}, \text{ or } \varphi = \int \omega dt,$$

in operator form

$$\varphi = \omega \frac{1}{S}.$$

Transient angular velocity has aperiodic kind at motor starting. By analogy to the general form of aperiodic process, the mathematical expression for transient angular velocity in operator form

$$\omega|S| = \frac{\omega_0}{T_m S + 1},$$

where ω_0 – synchronous angular velocity, c^{-1} ;

T_m – mechanical time constant.

Accordingly, the rotor angle in operator form

$$\varphi|S| = \frac{\omega_0}{T_m S^2 + S}.$$

Under the rules of transition from image to original form we received

$$\omega|t| = \omega_0 \left(1 - e^{-\frac{t}{T_m}} \right). \quad (3)$$

After integration we obtain the function of rotor angle time dependence

$$\varphi|t| = \omega_0 \left(t - T_m + T_m \cdot e^{-\frac{t}{T_m}} \right). \quad (4)$$

Using (4) we can get the time dependence of inductance $L(t)$.

Was introduced basic sinusoidal function

$$L_p(t) = L_{\sin} - A_{\sin} \cos \left(\varphi(t) \frac{2\pi}{\varphi_r} - \varphi_{dev} \right), \quad (5)$$

where L_{\sin} – middle of sinusoidal function;

A_{\sin} – amplitude of sinusoidal function;

$\varphi(t) \frac{2\pi}{\varphi_r}$ – angular velocity of sinusoidal function;

φ_{dev} – the initial angle of rotor position relative to the stator tooth.

Function (5) limited to levels of maximum and minimum inductance value, describes the inductance $L(t)$:

$$L(t) = \begin{cases} L_{\min}, & L_p(t) \leq L_{\min}; \\ L_p(t), & L_{\min} \leq L_p(t) \leq L_{\max}; \\ L_{\max}, & L_{\max} \leq L_p(t). \end{cases} \quad (6)$$

Then the transition phase inductance has form as shown on fig. 5.

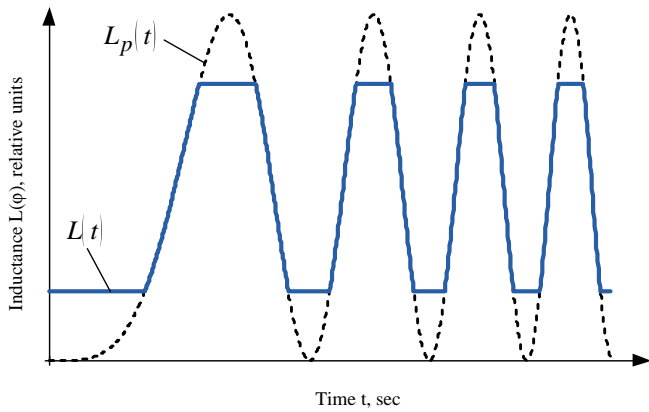


Fig. 6. Transition phase inductance.

Received function $L(t)$ was introduced in phase mathematical model of SRM. This model described by a system of differential equations:

$$\begin{cases} U_k \mathbf{C} = I_k \mathbf{C} R_k + L_k(t) \frac{dI_k \mathbf{C}}{dt} + \omega \mathbf{C} \Phi \mathbf{C} \\ J \frac{d\omega \mathbf{C}}{dt} = \sum_{i=1}^k M_k \mathbf{C} - M_c \mathbf{C} \\ \frac{d\omega \mathbf{C}}{dt} = \varphi \mathbf{C} \end{cases} \quad (7)$$

where $k = 1 \dots m$ – phase number;

$U_k \mathbf{C}$ – supply voltage of k -th phase;

$I_k \mathbf{C}$ – current of k -th phase;

R_k – ohmic resistance of k -th phase;

$\omega \mathbf{C}$ – angular velocity of motor;

$\Phi \mathbf{C}$ – total magnetic flux, $\Phi \mathbf{C} = \Phi_0 + \Phi_f \mathbf{C}$;

Φ_0 – base magnetic flux dependence on residual polarization;

$\Phi_f \mathbf{C}$ – phase magnetic flux dependence on phase current and inductance;

J – inertia moment of motor;

$\sum_{i=1}^k M_k \mathbf{C}$ – total electromagnetic torque of motor;

M_c – drag torque.

Block diagram model of one phase is shown on fig. 7.

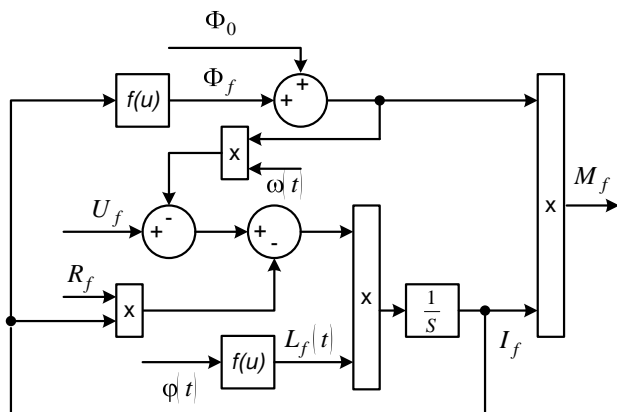


Fig. 7. Block diagram model of one phase SRM.

Research conducted for SRM with the following rated data:

- number of stator tooth, 6;
- number of rotor tooth, 4;
- power output, 190 kW;
- phase supply voltage, 560 V, U_r ;
- phase current, 155 A;
- angular velocity, 77.493 rad/sec;
- efficiency, 96 %;
- inertia moment, 8.68 kg·m²;
- ohmic phase resistance, 0.977 Ohm.

Simulation results are presented in fig. 8.

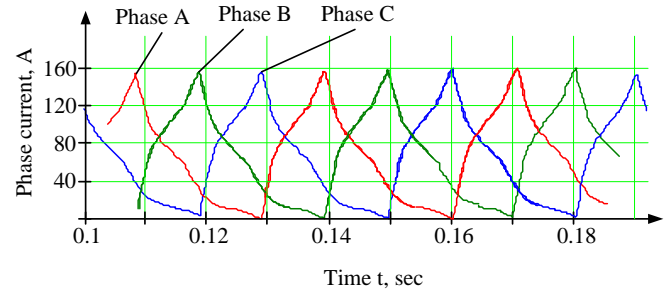


Fig. 8. Phase currents in steady state.

IV. CONCLUSION

The mathematical model of switched reluctance motor with variable phase inductance time dependence were received. On this model we can estimate inductance effect on motor electromagnetic process.

Simulation results improved that variable inductance has a significant effect on the current curve form. It permits to increase the mathematical model accordance to physical processes in the switched reluctance machine.

REFERENCES

- [1] DiRenzo, Michael T. Switched Reluctance Motor Control – Basic Operation and Examples Using the TMS320F240 // Texas Instruments Incorporated. Application Report, SPRA420A. – February, 2000. – 62 p.
- [2] Miller, T.J.E. Switched Reluctance Motors and their Control // Magna Physics & Clarendon Press. – Oxford, 1993. – 48 p.
- [3] Duijsen, P.J. van. Multilevel Modeling and Simulation of a Switched Reluctance Machine // Simulation Research. – The Netherlands, 2001. – 8 p.
- [4] Adaptive algorithm for SRM drives torque optimization / A. Argeseanu, O. Cornea, I. Torac // International Scientific Conference, eRA-2: The Contribution of Information Technology to Science, Economy, Society and Education. – Athens, Greece, 2007. – 8 p.

Energy output of photovoltaics and wind energy in Košice region during winter months

¹Martin BAČKO (3rd year), ²Oksana BRATASH, ³Vita OGAR
Supervisor : ⁴Dobroslav KOVÁČ

^{1,4}Dept. of Theoretical Electrotechnics and Electrical Measurement, FEI TU of Košice, Slovak Republic

^{2,3} Institute of Electromechanics, Energy Saving and Control Systems, Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine

martin.backo@tuke.sk, dobroslav.kovac@tuke.sk

Abstract—The paper deals with measurement of renewable energy sources – photovoltaic panel and wind turbine and possible utilization of their power to power up the heating system during winter months in Košice region. In first part the photovoltaics is explained, as well as placement of the module to get the most power. Second part deals with measured data and the results and third part deals with wind energy, shows the power curve of the turbine and compares it to measured data of wind speeds in region. At the end conclusion with some final thoughts for the future can be found.

Keywords—photovoltaics, wind energy, power, wind speed, efficiency

I. INTRODUCTION

In present days we are observing how fragile and unstable our current energy system is. It is based on fossil fuels as a primary energy sources, which are limited and their price per unit rises. This price increase of raw material leads to more expensive services, transport costs, and mostly more expensive heat and electrical energy. Because the demand for energy will rise, as the world's population is increasing, the current system is not possible to maintain forever. Alternative is needed and such alternative is possible with renewable energies like wind, sun and geothermal energy, which are available to whole world and are free. Their mass utilization proves to be imperative when the fossil fuels will be depleted and energy demand would need to be covered. Renewable (or „green“) energy had become extremely popular and hot topic in past few years. The technology evolved rapidly which lead for much higher affordability by general public. It is worth noting that although this energy can be utilized everywhere, the question is not if it can be used, but if it can be used effectively.

II. PHOTOVOLTAICS

When analyzing the utilization of solar energy by photovoltaic conversion, it is essential to know the solar map of the region, in other words the amount of solar energy to the unit of area (kWh/m^2). Such map for Europe and its countries was created by European Commission Joint Research Centre by Institute of Energy and Renewable Energy. The data was collected for years 2004-2010, so the data is pretty actual. The map (fig.1) represents average annual sum of solar energy for Slovak Republic. The average value is 1100-1200 kWh/m^2

so the region is suitable for utilization of photovoltaic energy. As can be seen from the map, northern mountain regions are less suitable, but the southern lowlands regions of Východoslovenská and Podunajská nížina have higher annual sum of 1300 kWh/m^2 provide better ground for photovoltaic modules. In average these areas have 2000-2200 sunny hours per year (the theoretical maximum is 4447 hours).

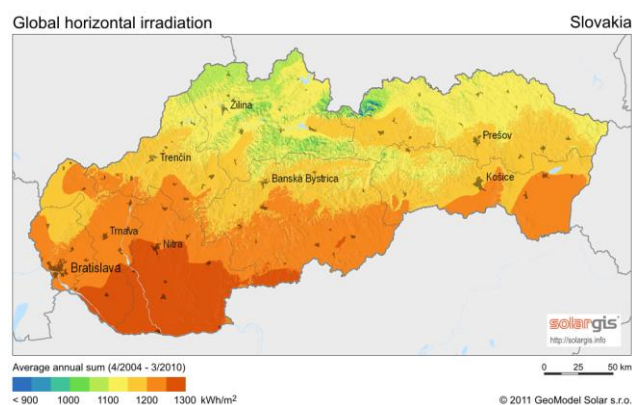


Fig.1 Solar map of Slovak republic

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 the 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.2).

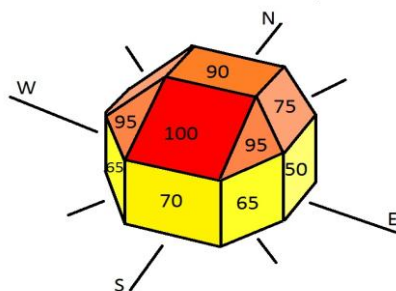


Fig.2 Percentage of optimal energy production at different orientations and angles

In order to better photovoltaic understanding, the V-A (fig.3) characteristics of module are needed to be explained a bit.

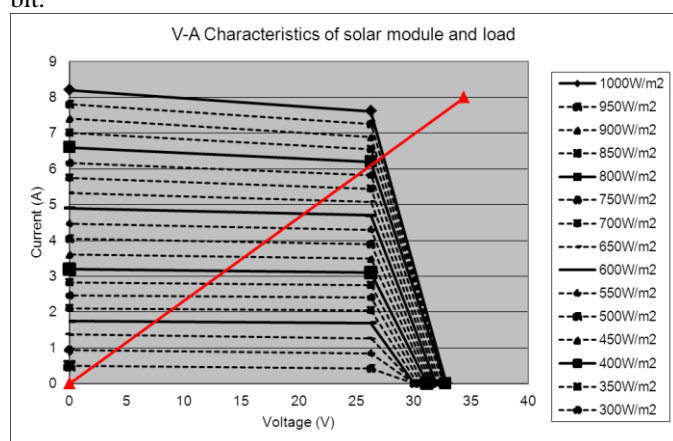


Fig.3 V-A characteristics of photovoltaic module

This V-A characteristics clearly shows that the maximum power is directly proportional to the intensity of sunlight. When the intensity decreases, the maximum obtainable power decreases as well. It is worth noting the few key values for every datasheet. The rated power for the module which was measured was P_{max} at 200W. Voltage at P_{max} is 26,3 V. This voltage is referred as V_{mp} . The V-A characteristics shows that the V_{mp} value is the same regardless of sunlight intensity. This is the value of voltage when the maximum power can be obtained. Current at P_{max} is 7,61A. This current is referred as I_{mp} . Short-circuit current (I_{sc}) is 8,12 A and open-circuit voltage (V_{oc}) is 33,4 V for intensity of 1000W/m² of sunlight. The red line represents the load which was used for measurement. The load was resistor with the value of 4,3 Ω , therefore the linear line. The results from measurements were only values on the red line, however as can be seen from the V-A characteristics (fig.2) with properly dimensioned load, the maximum power for every value of intensity can be obtained.

III. MEASUREMENTS AND RESULTS

Program EasyLog was used for measurement of voltage. This voltage was measured every day in minute intervals. With such intervals 1440 values of voltage per every day of year were measured and stored to file. Because the value was known, the current was calculated using Ohm's law:

$$I = U/R \quad (1)$$

With the voltage and load known, the obtained power could be easily calculated using the equation for power in DC circuits which is:

$$P = U^2/R \quad (2)$$

As can be seen from the V-A characteristics (fig.2), especially in case of lower intensities, the values of current and power were quite low, however this was caused by the load which was chosen. It is important to look at P_{max} values at V_{mp} (26,3V) because these are the value which can be obtained.

With the known currents, next step was the determination of maximum obtainable power. As can be seen from the picture, although the power can be calculated as:

$$P = U \cdot I \quad (3)$$

In order to obtain the real values the calculated current has to be lowered by a specified amount, because the line is not straight, but with increased voltage it is decreasing. This decrease is more significant in higher intensities, as can be seen.

This decrease of current for V_{mp} voltage was calculated as:

$$K = I_{mp}/I_{sc} \quad (4)$$

For the intensity of 1000W/m² this decrease is $K = 7,61/8,12 = 0,94$.

Therefore the maximum obtainable power is calculated as:

$$P_{max} = V_{mp} \cdot I \cdot 0,94 \quad (5)$$

Upper mentioned equation (5) is valid for currents from range 6,6 to 8,12 A, which are for range band 800-1000 W/m². The same procedures were made for every remaining range bands (800-600 W/m², 600-400 W/m², 400-200W/m², 200-0 W/m²).

The level of decrease is different in every range band, in the last 200-0 W/m² the coefficient of decrease was only 0,97.

One of the goals of the work is to evaluate the performance of photovoltaic panel especially during winter months, as the question is, whether it can provide enough energy for domestic heating system, which can be for example electrical cauldron or thermal heat pump, which is much better option, despite higher initial investment.

Figure (fig.4) represents the graphical representation of average daily values of obtained power and obtainable power P_{max} .

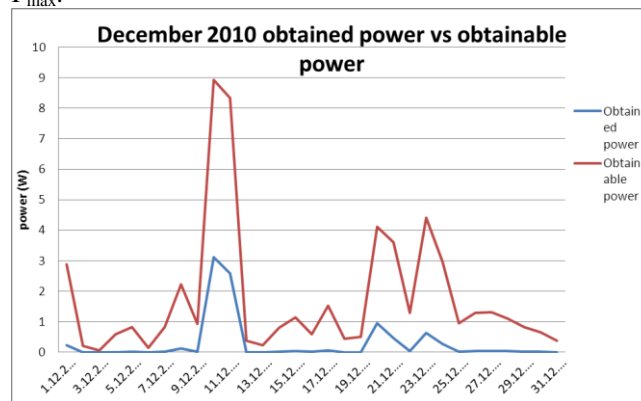


Fig.4 December 2010 power graph

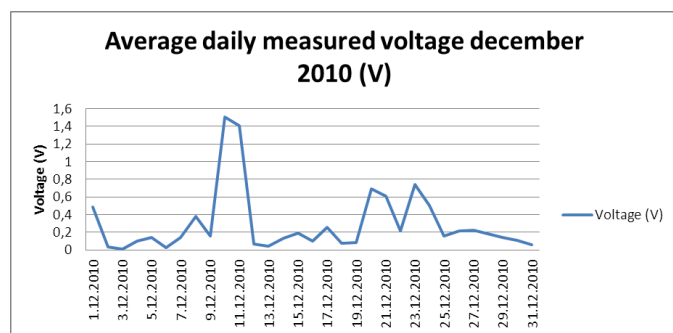


Fig.5 December 2010 average daily voltage

The graphs show clearly that in the month of December, when the amount of solar energy is lowest in year, the photovoltaic panel measured in Košice region is absolutely useless. However it is fair to note, that this statement is completely true only for December 2010, because the weather cannot be exactly predicted, December 2011 was different, weather in December 2012 can be/won't be different.

Figures (fig.6), (fig 7) show the results for January 2011.

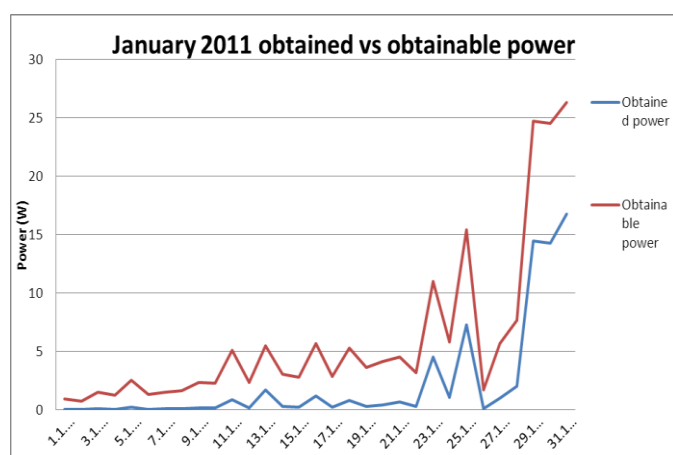


Fig.6 January 2011 power graph

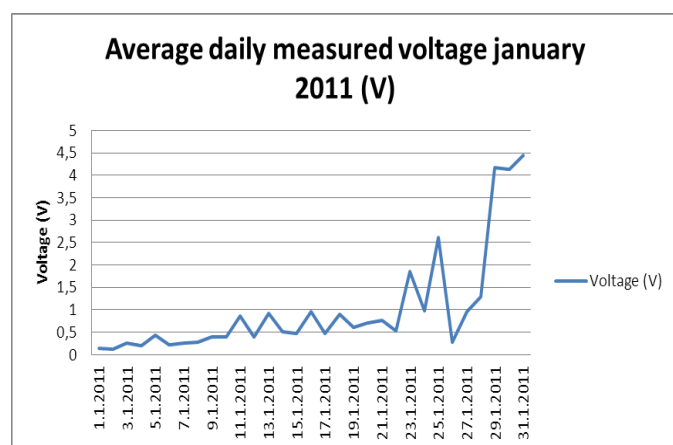


Fig.7 January 2011 average daily voltage

The situation in January is getting better towards the end of the month, however if 200W rated panel for 300€ can provide only 15 watts of power (which cannot be granted because are subject to weather), the result is unsatisfactory. The measurements prove that especially during winter months photovoltaic is unreliable source of energy and cannot be considered as a primary source in the Košice region.

IV. WIND ENERGY

Wind energy can be very effective source of energy when the right conditions are met. Small wind turbines with diameter about 2 meters can provide 500-1500 W of power for a very reasonable price, comparable with photovoltaic panels. These turbines are small enough to be mounted on houses and widely used. They generate alternating current and need charging regulator to prevent from overcharging of batteries.

In order to be effective they need to be mounted high, at least 10 meters above ground and more importantly, they need the specified wind speed to operate effectively. For better understanding, the power curve of the typical (HAWT – horizontal axis wind turbine) turbine is pictured in figure (fig.8).

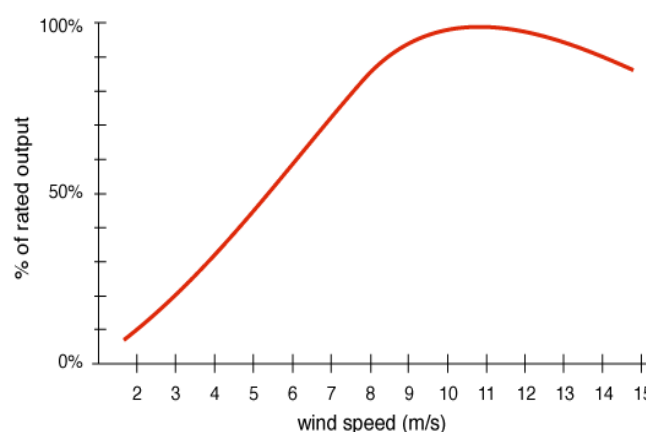


Fig.8 Power curve of typical HAWT turbine

For most turbines the cut-in speed which is the speed when wind turbine begins to generate energy is 2-3 m/s. The nominal wind speed, when the turbine reaches its nominal power output is mostly 10-14 m/s based upon turbine type. Shut off speed is about 18-20 m/s and more, this is speed, when there is danger of physical damage, therefore the turbine locks itself using the brakes.

Figure (Fig.8) clearly shows that in order to get at least 50% power from the turbine, constant wind speed at least 5-6 m/s is needed. This wind speed is common in coastal areas where the turbines are mostly used.

Figure (Fig.9) shows the average monthly wind speed in Košice city in years 2009-2011.

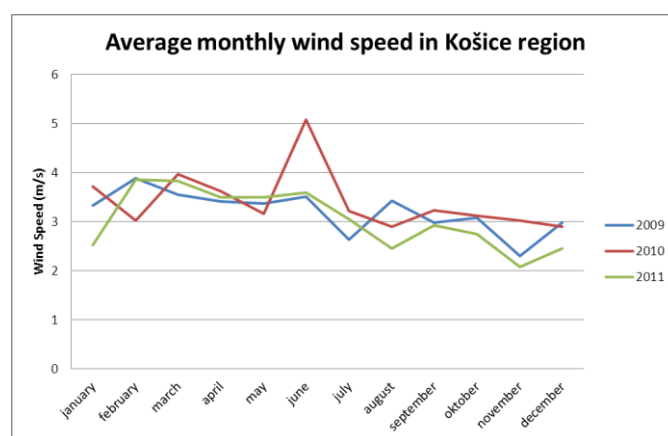


Fig.9 Average monthly wind speed in Košice

The graph (fig.9) clearly shows that the monthly average wind speed is about only 3 m/s which is barely the cut-in

speed for most of the small HAWT turbines. With such speed, maybe 10-15% of nominal power is obtained, which is again very low to consider as a primary source of electric power. Unfortunately only few areas of Slovak Republic are suitable for utilizing wind energy, where the wind energy reaches high speeds constantly. Another problem, which prevents the wider usage in cities is the amount of wind blasts between block of flats, turbulences and wind tunnels because of which the turbine must be mounted high, which is often not possible.

V. CONCLUSION

The results of measurements indicate that in current state of technology, the renewable energy sources are not suitable for utilization in Košice region. The wind energy would be interesting to consider in case, that the new turbine with lower nominal wind speed requirement can be manufactured. Measurement of photovoltaic panel proves that the panels are more suitable for different region with where the Sun's intensity is higher. On the other hand, the panels were considered for the winter months, which are mostly cloudy, therefore low power output is expected. During the summer months, the panel provided 70-80% of its nominal power, therefore the question of accumulating energy and its proper storage for utilization during winter months is still valid and worth considering.

ACKNOWLEDGMENT

The paper has been prepared under support of Slovak grant projects KEGA No. 005TUKE-4/2012, KEGA No. 024TUKE-4/2012 and VEGA No. 1/0559/12.

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, 2006 Netherlands,
- [3] Eicker Ursula: Solar Technologies for Buildings, John Wiley & Sons, 2003, England
- [4] I. Kováčová, D. Kováč, T. Vince, "Electromagnetic compatibility", Published by TU FEI, Košice, pages 138, 2009, ISBN 978-80-553-0150-1
- [5] D. Kováč, I. Kováčová, J. Molnár, "Elektromagnetic compatibility – measurement", Published by TU, FEI, pp. 72, 2009, ISBN 978-80-553-0151-8
- [6] I. Kováčová, D. Kováč, D. "Industrial Electrical Engineering – Modeling of Converters", 1st part, book, Published by TU of Košice, 2011, 66 pages, ISBN 978-80-553-0617-9
- [7] I. Kováčová, D. Kováč, D. "Industrial Electrical Engineering – Modeling of Converters", 2nd, part, book, Publisher TU of Košice, 2011, 68 pages, ISBN 978-80-553-0618-6

Frequency regulation with energy storage

¹Tomáš Košícký (2nd year)

¹Department of Electric Power Engineering , FEI TU of Košice, Slovak Republic

¹tomas.kosicky@tuke.sk

Abstract—The possibility of dynamic energy storage as an add-on function of FACTS regulators enables the control of both active and reactive power at the point of connection . Dynamic control of active as well as reactive power runs independently of each other. By control of the reactive power , grid voltage and stability are safeguarded with high dynamic response. By control of the active power, By control of active power, new services based on dynamic energy storage may be added. Integration of dynamic energy storage into transmission and distribution systems has the potential to provide significant benefits to the supply chain. Increasing penetration of distributed generation, particularly based on renewable energy sources, is driving the need for distributed energy storage to provide supporting services that will allow existing networks to deliver reliable, cost effective, environmentally attractive, and high quality electricity.

Keywords—Battery storage, Power systems, SVC, STATCOM

I. INTRODUCTION

A series of coordinated contingency reserves are maintained to deal with sudden, unexpected loss of electric power generation or transmission, or with rapidly increasing load. Primary, secondary and tertiary regulation reserves are available gradually according to the time disposition of the reserve. The fastest services of the primary reserve, also called Primary Frequency Regulation come from generators that are on-line, spinning, and loaded at less than full output. Supplemental reserves, also called Secondary Frequency Regulation [1], shall be fully available within 15 minutes but could come from fast-start hydrogenerators [2]. Cold reserve, also called Tertiary Frequency Regulation can come from slower resources and can be available within 30 minutes [2].

In case of load-generation imbalance, Primary frequency Regulation requires prompt spinning reserve to provide active power for durations up to 30 minutes [3]. Battery systems should absorb or provide energy as it fluctuates.

Energy storage systems equipped with fast-acting grid interface power electronics offer an alternative to the traditional strategy of maintaining adequate spinning reserve margin to mitigate frequency contingencies. In response to frequency contingencies, energy storage systems can supply “prompt” spinning reserve i.e., rated power deployed within a few cycles for a sufficient period to enable other generation assets (secondary regulation) to be brought on line. In

addition, energy storage can help the operator reduce or eliminate the need for purchase of supplemental services of primary regulation provided by generators. This approach avoids the capital and operating costs associated with continuously operating spinning reserve generation.

The FACTS system with Energy Storage, combines dynamic energy storage provided by high voltage Saft Li-ion batteries with SVC Light for reactive power compensation and dynamic voltage control. The Voltage Source Converter (VSC) technology such as SVC Light has the advantage of being able to almost instantly change its operating point within its four-quadrant power area if an energy source is included. Thus, SVC Light with Energy

Storage can modulate both reactive power, as an ordinary SVC Light, as well as active power thanks to the batteries. This can be used to support the grid with the best mix of active and reactive power during stressed conditions.

II. DESIGN OF AN ENERGY STORAGE FOR FREQUENCY REGULATION

As already known, a STATCOM is a power electronic converter used as a reactive power compensator. With a STATCOM, the reactive power can be varied continuously. In addition, the speed of response of STATCOM is higher than traditional SVC systems, thereby being able to counteract much faster variations, e.g. voltage, by varying the reactive power. STATCOM utilizes a SVC connected in shunt to the grid at both distribution and transmission level.

The energy storage, which can be added to SVC Light, is based on string of batteries, since a high amount of energy is needed and the required discharge time is in the range of minutes to hours [4,5]. Figure 1 illustrates system of strings of batteries consisting of an SVC Light together with a number of parallel-connected battery strings on the DC-link in order to obtain high power. Each battery string consists of a number of series-connected batteries in order to build up the required voltage level. If the minimum requirement is to size the battery system to supply 5 MW for 30 minutes, consequently the minimum energy capacity of the battery system is 2.5 MWh. However, since the energy storage level of the energy storage

is kept at around 50 %, (state of charge - SOC) the energy storage is then designed for an energy capacity of 5 MWh.

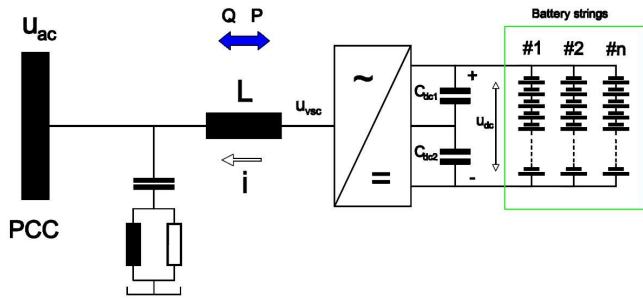


Fig. 1. Schematic layout of SVC Light with Energy Storage

Battery cell has a discharge power of 146 W per cell. Thus, for an energy capacity of 5 MWh, the minimum number of battery cells required is 34247.

Parallel strings of batteries provide rated DC voltage on the converter and current required for the discharge time. Then there are 34247 installed battery cells divided into 2 strings. With this optimal design, the installed energy capacity of the battery system is 5.36 MWh. It is possible to design an energy-storage system with the minimum amount of energy for dedicated operation in regulation service only. However, such a single-use system (regulation use) can have a number of disadvantages. From a grid-operation standpoint a storage system with limited energy may require frequent intervention at significant power levels in order to prevent the storage from becoming too full or too empty. A methodology for maintaining the battery system within a desired SOC band is discussed in the next section.

From a battery standpoint limiting the energy rating has two main impacts. First, each charge-discharge cycle will represent a larger percentage of the installed energy, causing the available cycle life to be used up more quickly. Second, the power-to-energy ratio for charging may be too high to be compatible with a long battery life. For conventional Li-ion batteries with graphite negative electrodes, charging at very high rates can result in a significant level of premature aging, due to a process of lithium plating. The acceptable charge power is a function of the battery's performance type (e.g. high power, medium power or high energy) and of the installed energy. Another factor is that energy-storage systems can often be more easily justified if they can address multiple value streams. Extending it with additional battery, system provides synchronized reserves. Sizing the battery system to supply 5 MW for 30 minutes, and using it to supply ± 5 MW of regulation service, will have the following benefits:

- It allows both regulation and reserve services to be supplied by the same system.
- The average cycle depth is reduced, thus extending the operating life.
- The one-hour discharge time allows use of a more cost effective medium-power battery option (rather than highpower), while still providing 5 MW of charge power without serious life impact.

III. POWER-FREQUENCY CHARACTERISTIC OF THE PRIMARY FREQUENCY CONTROLLER

Figure 2 illustrates the principle of frequency profile (left) and power-frequency characteristic of the primary frequency regulation (right). When the frequency is within the deadband, the power from the battery is zero. When the frequency gets outside the deadband, the battery charges/discharges. If the frequency is above the upper limit of the deadband, then the battery charges. If the frequency is below the lower limit of the deadband, then the battery discharges.

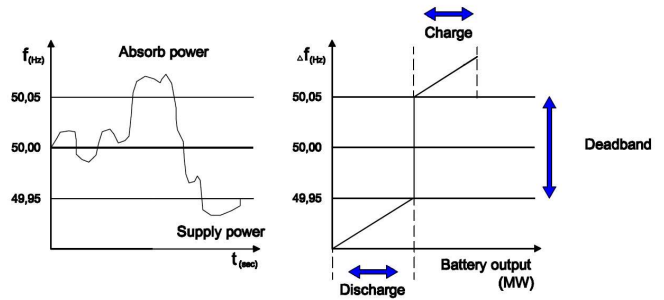


Fig. 2. Operation principle

Since the frequency deviation with respect to the fundamental frequency can be either positive or negative, the battery storage should be in a state allowing both discharge (negative frequency deviation) as well as charge (positive frequency deviation), at any time. Therefore, the average SOC level, of the battery storage is chosen to be 50 %. On the other hand, the issue is to keep the SOC level around 50 %. The reason is that, after a discharge cycle, other consecutive discharge cycles might possibly occur and the battery storage might end up fully discharged; the SOC level is below 0. Similarly, after a charge cycle, other consecutive charge cycles might possibly occur and the battery storage might end up fully charged; the SOC level is above 100 %.

Note that the short cuts in the corners seen in Fig. 4 are simply caused by the fact that the frequency did not go through the corners all the time. One solution to this issue is to have a separate power to charge or discharge the battery storage when the frequency is within the assigned deadband and when the SOC level is outside a certain tolerance band, which is assumed to be 45 % to 55 % in this case.

As can be seen in Fig. 4, when the frequency is within the deadband, the power from the battery is no longer zero but a value corresponding to the SOC level of the battery. When the frequency is within the deadband and if the SOC level of the battery is above a certain maximum value, which is 55 % in this case, then the battery gets discharged.

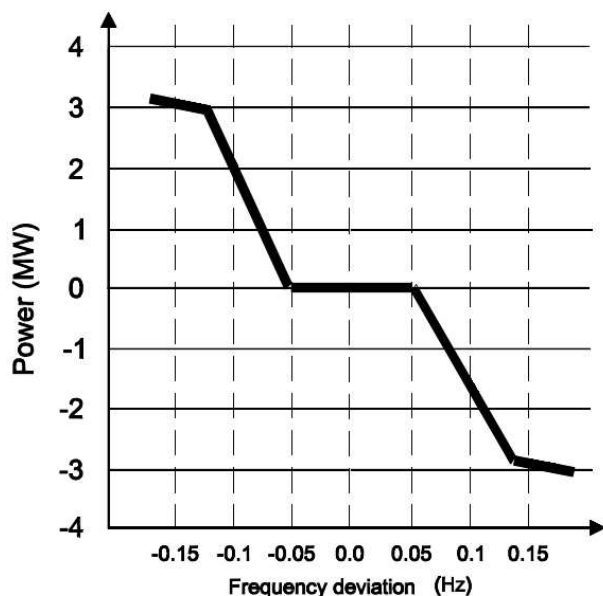


Fig. 3. Power-frequency characteristic

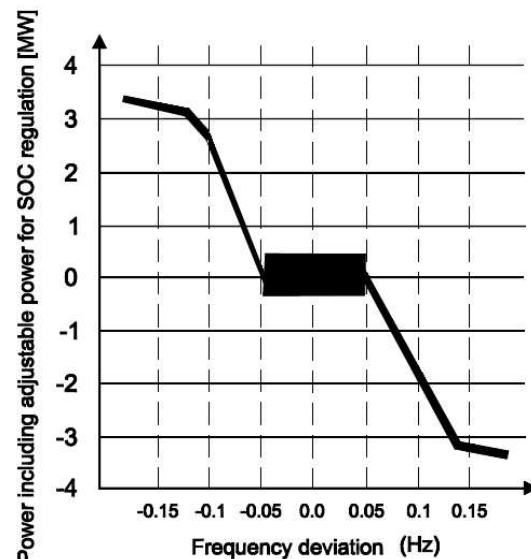


Fig. 5. Power-frequency characteristic having a SOC control loop

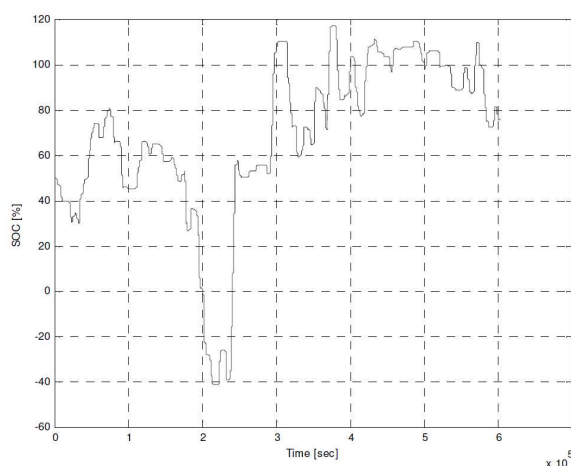


Fig. 4. SOC level as a function of time

Similarly, when the frequency is within the deadband and if the SOC level of the battery is below a certain minimum value, which is 45 % in this case, then the battery gets charged. In other words, according to the SOC level, the line of zero power moves upward and downward in a continuous manner giving rise to the black area in Fig. 5.

The SOC levels lie between 5 % and 95 %, which is acceptable from the point of view cycling of the battery. In addition, it can be concluded that the battery storage of ± 5 MW is enough for the frequency regulation in area considered by our simulation.

IV. IMPLEMENTATION OF THE SOC CONTROL LOOP

Figure 6 shows the implemented SOC control, where f is the grid frequency, P is the power according to the characteristics in Fig. 3 and SOC is the SOC level.

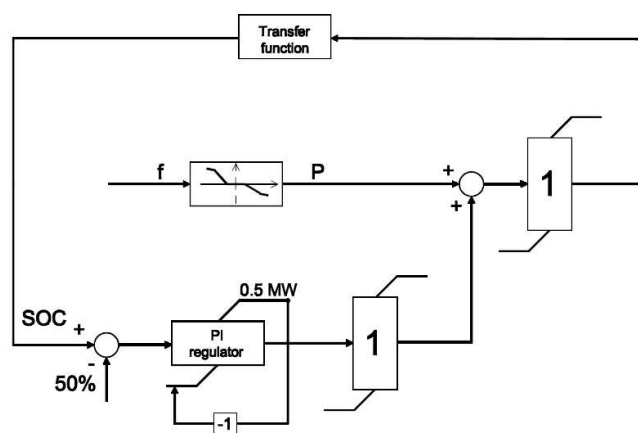


Fig. 6. SOC level as a function of time having a SOC control loop

To evaluate these larger SOC swings, it is necessary to filter out (i.e. ignore) smaller capacity fluctuations. The level of filtering will obviously affect the results; taking two extremes, it can be seen that no filtering will yield the same average results as the microcycle analysis, while filtering out all SOC fluctuations of less than 80 % will yield virtually no cycles. For the purposes of this evaluation a filtering level of 10 % is used, i.e. any SOC swings of less than 10 % are ignored by the analysis.

V. CYCLE LIFE DATA OF A LITHIUM ION BATTERY

Batteries degrade in predictable ways when subjected to repeated charge/discharge cycles in a controlled manner. Figure 7 shows a cycle-life curve for the Saft lithium-ion battery system based on depth of discharge for each cycle [6,7].

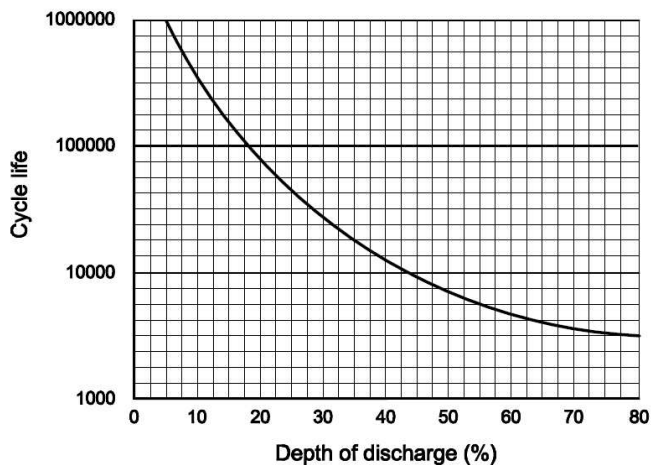


Fig. 7. Cycle-life curve for the Saft lithium-ion battery system

VI. CONCLUSIONS

From this analysis, it can be seen that battery energy storage is an effective means for providing primary frequency regulation. The implementation of a frequency deadband and a SOC control loop maintains the battery within a reasonable SOC band and minimizes the impact of the cycling operation on battery life.

VII. REFERENCES

- [1] B. J. Kirby, "Frequency Regulation Basics and Trends," DOE report TM-2004/291, 2004.
- [2] EPRI-DOE Handbook, "Energy Storage for Transmission and Distribution Applications", Dec 2003.
- [3] A. Nourai. (2003, Dec.). Comparison of the Costs of Energy Storage Technologies for T&D Applications [Online]. Available: <http://electricitystorage.org/pubs/2004/EPRIDOE%20Storage%20Costs-ESA.pdf>
- [4] J. Svensson and W. Hermansson, "Dynamic Energy Storage using SVC Light," Cigré conference Paris, France, August 24-19, 2008.
- [6] Proceedings of Electrical Energy Storage Applications and Technologies (EESAT) Oct 2009, Seattle.
- [7] IEEE Draft Recommended Practice for the Characterization and Evaluation of Emerging Energy Storage Technologies in Stationary Applications, IEEE Standard P1679TM/D13-2010, Jan. 2010.

Harmonic Potential Field Method for Path Planning of Mobile Robot

¹Michal KALAVSKÝ (3rd year),
Supervisor: ²Želmíra FERKOVÁ

^{1,2}Dept. of of Electrical Drivers and Mechatronics, FEI TU of Košice, Slovak Republic

¹michal.kalavsky@tuke.sk, ²zelmira.ferkova@tuke.sk

Abstract—Harmonic potential field method is one of the ways how to eliminate the weakness of potential field method for path planning of mobile robot. This method is more convenient modification of potential field method. Principle of this method consists in creation of global minimum of a virtual space that is represented a real space. Global minimum is the goal position so no local minimum exists. In this paper, simulation results of this method will be used for static system. It means that position of obstacles will be fixed.

Keywords—potential field, harmonic function, boundary conditions

I. INTRODUCTION

In recent years known are several methods for path planning of mobile robot. What does planning mean? Planning has several definitions but for us important is the one used in robotics. So we can define planning as finding an algorithm that is able to satisfy safe path from start to the goal position without colliding with obstacles. Finding this algorithm is quite difficult and time consuming. Thus, finding this algorithm with use of harmonic potential field method is subject of my work and therefore, in this paper, described will be the harmonic potential field method.

Harmonic potential field method pertains to planning on grid. It means that real space is represented by grid. Imagine real space where robot is moving. We can divide it into cells. Every cell has his evaluation. By this process achieved is a virtual space. This virtual space is represented by the grid with finite number of cells. So motion of robot is controlled by evaluation of cells.

Path planning using potential field method consists of two problems that are necessarily to solve. The first problem is generation of potential field that is represented real space. And the second problem is to find algorithm that is capable find safe path in potential field.

In this paper will be described the finding of algorithm for safe path in the created harmonic potential field. In the first section of this paper potential, potential field and harmonic function will be described, in the second harmonic potential field and in the third experimental results.

II. POTENTIAL FIELD AND HARMONIC FUNCTION

Consider vector field F . We can define potential of this field as follows: The potential ϕ of vector field F can be defined as the work function or as its negative in form:

$$\vec{F} = \nabla \phi \quad \text{or} \quad \vec{F} = -\nabla \phi \quad (1)$$

In accordance to [1] potential ϕ of vector field F is defined as follows: If particles of like sign attract each other (gravity fields), then $\vec{F} = \nabla \phi$ and the potential is equal to work made by the field. If particles of like sign repel each other (electrostatic fields), then $\vec{F} = -\nabla \phi$ and the potential is equal the work made against the field by the particle.

Now consider conservative vector field F . Conservative field F is defined in accordance to [1] as scalar potential ϕ that has form $\vec{F} = \nabla \phi$ or $\vec{F} = -\nabla \phi$. If $\vec{F} = \nabla \phi$ and the vector field F is conservative, then the vector field F is termed as potential field. Potential field has his properties. One of them is described as follows: The potential ϕ of the vector field F , under certain conditions, satisfies an important second-order differential equation called Laplace's equation:

$$\nabla^2 \phi = 0, \quad (2)$$

at points not occupied by sources of F .

Consider two-dimensional case. Laplace's equation (2) can be described in form:

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0 \quad (3)$$

From previous statements we can define a harmonic function as follows: Harmonic function is function that satisfies the following criteria:

1. Satisfy Laplace's equation $\nabla^2 \phi = 0$.
2. Have continuous, single-valued first derivatives.
3. Have second derivatives.

From previous declaration we can say that harmonic function throughout a region R must have maximum and minimum on the boundary of region R and none within R itself. This is very useful for harmonic potential field method because this method must have only global minimum and maximum and none local minimum and maximum. Harmonic function is exactly basis of the harmonic potential field method.

The definition of the second derivative of harmonic function demonstrates very important property of a harmonic

function. For one-dimensional case, the second derivative of harmonic function $\varphi(x)$ is given as follows:

$$\varphi(x) = \frac{1}{2} \lim_{\Delta x \rightarrow 0} [\varphi(x - \Delta x) + \varphi(x + \Delta x)] \quad (4)$$

From equation (4) results that the value of a harmonic function φ at any point is the average of φ at its neighboring points. This statement is suggested familiar property of a potential: A function can have no minimum or maximum within a region in which it is harmonic.

Find a solution of Laplace's equation is solving boundary values problem (Dirichlet type). In this case, it is needed to find formulation φ for entire region R where within region R must satisfy Laplace's equation $\nabla^2 \varphi = 0$ and specific values are given for boundary of region R . This case is equal as harmonic potential field method case where it is needed to create a harmonic potential field with minimum and maximum on its boundary and within region R it satisfies Laplace's equation $\nabla^2 \varphi = 0$. [1] [6]

III. HARMONIC POTENTIAL FIELD METHOD

Harmonic potential field is one of the ways how weakness of potential field eliminates. Purpose of using harmonic potential field is to eliminate creation of local minimum in virtual space. The basic idea of harmonic potential field arises from harmonic function above described.

We can define harmonic potential field with accordance to [2], [3] as harmonic functions that are solutions of Laplace's equation. From this results that if function is harmonic, then must satisfy Laplace's equation on the region $\Omega \subset R^n$ in form:

$$\nabla^2 \varphi = \sum_{i=1}^n \frac{\partial^2 \varphi}{\partial x_i^2} = 0 \quad (5)$$

where x_i represent Cartesian coordinates.

Now consider equation (5) for path planning of mobile robot when it is needed to create virtual space (harmonic potential field). Now the question arises how to define equation (5) for every part of region Ω ? Harmonic potential field method is based on principle of minimum and maximum. It means that motion of robot is controlled from the highest to the lowest value in a virtual space. Therefore, it is necessary to define maximum and minimum values in the virtual space. Exactly harmonic function offers this property. If function is harmonic on the region Ω , then the function has minimum and maximum only on boundary of the region Ω . No local minimum and maximum exists within the region Ω . And this is very important property of harmonic function for harmonic potential field method. Due to this property using of harmonic function is convenient for creation of harmonic potential field without local minimum.

Now consider two-dimensional region $\Omega \subset R^2$. In this case Laplace's equation has form:

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = 0 \quad (6)$$

Seeing that φ is harmonic function and satisfies Laplace's equation (19) so function φ has his minimum and maximum only on boundary of $\partial\Omega \subset \Omega$. Therefore, it is convenient for us to define boundary conditions as boundary of all obstacles and boundary of goal position. Then created harmonic

potential field will be satisfied the following boundary conditions of Dirichlet type in form:

1. $\varphi|_{\partial\Omega} = c$
2. $\varphi_{(goal)} = 0$

where c is constant.

The first condition says that boundary of all obstacles have maximum value on the region Ω . The second condition says that goal position has minimum value on the region Ω . If boundary conditions are defined by this way, then potential field is harmonic field with only global minimum represented by goal position. We can say that boundary conditions of region Ω are defined.

Now it is needed to define other parts of region Ω . For other parts of region Ω is used Laplace's equation (20). Solution of Laplace's equation is obtained by using methods of numerical mathematics as for example Gauss-Seidel method, Jacobi method or over-relaxation method.

Consider a virtual space that is represented a real space and let it is two-dimensional case R^2 . First the space is needed to divide on m -cells where every cell has his coordinates $[x_i, y_j]$. Let the distances between the cells are equal what means that we are creating a regular grid with a finite number of cells (see Fig. 1). Let this distances are denoted as parameter h in x and y directions. By this process the regular grid is created where every cell is representing a fixed part of the virtual space.

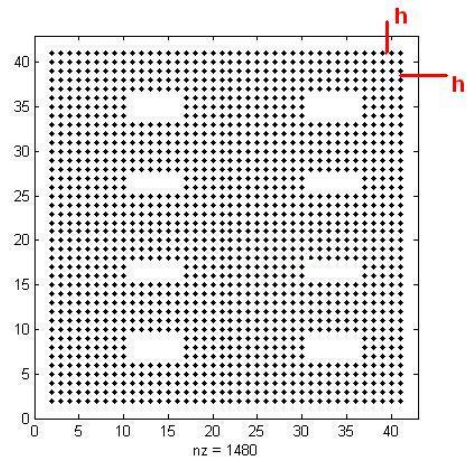


Fig. 1 Virtual regular grid represented real space

Let $\varphi(x_i, y_j)$ represent a discrete regular sampling of potential value in a grid and be a function which satisfies Laplace's equation. After dividing of a space to a grid it is necessary to find approximation of the partial differential equation at each grid point. In accordance to [5] a central difference formula for the second derivatives of φ can be derived using Taylor series expansion for x and y coordinates:

$$\frac{\partial^2 \varphi(x_i, y_j)}{\partial x^2} \approx \frac{\varphi(x_{i+1}, y_j) - 2\varphi(x_i, y_j) + \varphi(x_{i-1}, y_j))}{h^2} \quad (7)$$

$$\frac{\partial^2 \varphi(x_i, y_j)}{\partial y^2} \approx \frac{\varphi(x_i, y_{j+1}) - 2\varphi(x_i, y_j) + \varphi(x_i, y_{j-1}))}{h^2} \quad (8)$$

where h denotes the step size (distance between cells) in approximating the derivative in x and y directions. Now we can adopt the algebraic equation by (8) + (9):

$$h^2 \left(\frac{\partial^2 \varphi(x_i, y_j)}{\partial x^2} + \frac{\partial^2 \varphi(x_i, y_j)}{\partial y^2} \right) = \varphi(x_{i+1}, y_j) + \varphi(x_{i-1}, y_j) + \varphi(x_i, y_{j+1}) + \varphi(x_i, y_{j-1}) - 4\varphi(x_i, y_j) \quad (9)$$

Seeing that $\varphi(x_i, y_j)$ is harmonic function, then $\varphi(x_i, y_j)$ satisfies Laplace's equation as shown in (5) and (6). Therefore, $\varphi(x_i, y_j)$ can be calculated by equation:

$$\varphi(x_i, y_j) = \frac{1}{4} \cdot (\varphi(x_{i+1}, y_j) + \varphi(x_{i-1}, y_j) + \varphi(x_i, y_{j+1}) + \varphi(x_i, y_{j-1})) \quad (10)$$

By analogy with (4) we can say that from equation (10) results very important statement: Value of harmonic function φ at a point with coordinates $[x_i, y_j]$ is equal to the average of φ at its neighboring points.

Since the system described by (10) is usually quite large, methods of numerical mathematics mentioned above can usually be used. In this paper, we use the Gauss-Seidel method. By using Gauss-Seidel method can be obtained the final equation that describes potential field:

$$\varphi^{k+1}(x_i, y_j) = \frac{1}{4} \cdot (\varphi^k(x_{i+1}, y_j) + \varphi^{k+1}(x_{i-1}, y_j) + \varphi^k(x_i, y_{j+1}) + \varphi^{k+1}(x_i, y_{j-1})) \quad (11)$$

By equation (11) and sufficient number of iterations potential field is created. If the boundary conditions of the virtual space R^2 are satisfied (goal position is global minimum, boundary of all obstacles are global maximum), then robot's motion is controlled by tracking of the lowest value of this space. So robot is able to arrive to the goal position because is the lowest value of this space (global minimum). [1] [3] [4]

IV. EXPERIMENTAL RESULTS

As consider above, path planning using harmonic potential field method consists of two problems that are necessarily to solve. The first problem is generation of potential field that is represented real space. And the second problem is to find algorithm that is able to find safe path in the potential field.

The first problem can be solved by using software Student's Quick Field that is able to create harmonic potential field with only one global minimum.

Target of our study is to create the algorithm that is applicable at the harmonic potential field. It means that created algorithm is able to provide finding of safe path to the goal position from whichever position in the harmonic potential field. Algorithm and all simulation results were created in software MATLAB (version 7.6.0).

Consider the virtual space depicted on Fig. 2. You can see the space that was used for simulation of robot's motion. Created field has the dimension of 42x42 points. Free (white) space represents bound of obstacles and bounds of space. Every (black) dot represents space where robot can move and has its evaluation. By this evaluation of every point it is possible to find safe path to goal position. Shown in Fig. 2 is the image of goal position as green circle. The goal position has coordinates $[x=31, y=18]$. Forasmuch as harmonic potential field has only one minimum, global minimum, the goal position is global minimum and has the highest evaluation of this potential field.

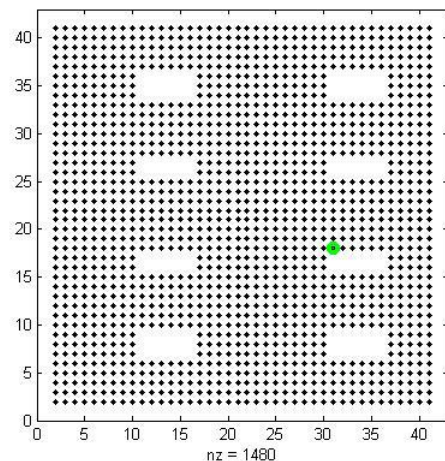


Fig. 2 Virtual space used for simulation of robot's motion

Algorithm for finding safe path in created field was made using software MATLAB. You can see created program for simulation of robot's motion in Fig. 2. The program draws field where depicted is simulation of robot's motion, four pushbutton and five fields of text. By fields "Put x-coordinate" and "Put y-coordinate" loaded are x- and y-coordinates. By pushbutton "Download coordinate" drawn are coordinates as a blue circle in simulation space. In this case entered were coordinates $[x=36, y=6]$. This is shown in Fig. 3.

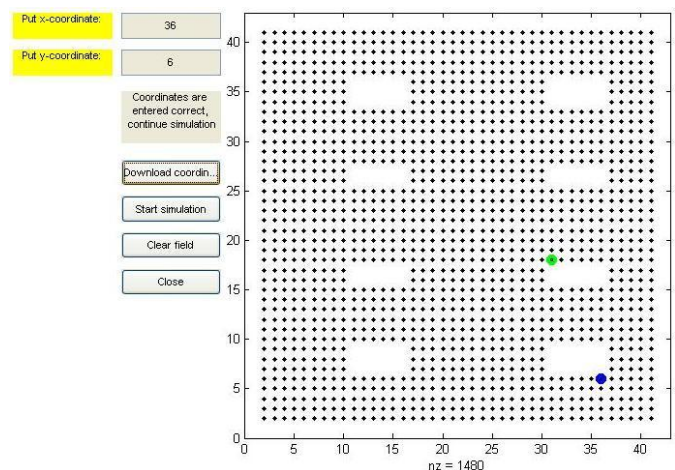


Fig. 3 Drawing of start position during simulation of robot's motion

If entered coordinates are incorrect (are in position of obstacles – white free space) it is needed to enter the coordinates again. If entered coordinates are correct simulation can continue. Entered coordinates are drawn as blue circle. Now we can start simulation by pushbutton "Start simulation". Result of simulation can be seen in Fig. 4.

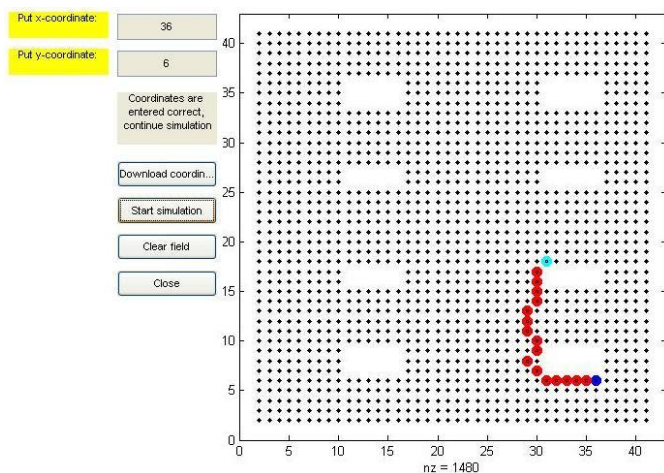


Fig. 4 Simulation result, final path from start to the goal position

The algorithm in created program is based on comparison of evaluation of elbow neighbours. Every point in created field has eight neighbours. So robot can move in any direction. Pushbutton “Clear field” clears the result of previous simulation and program is ready to put coordinates again and to start simulation. Pushbutton “Close” breaks program. Seen in Fig. 5 can be other sample of moving robot, using program created in MATLAB. Robot is able to come from whichever start position to goal position without colliding with obstacles.

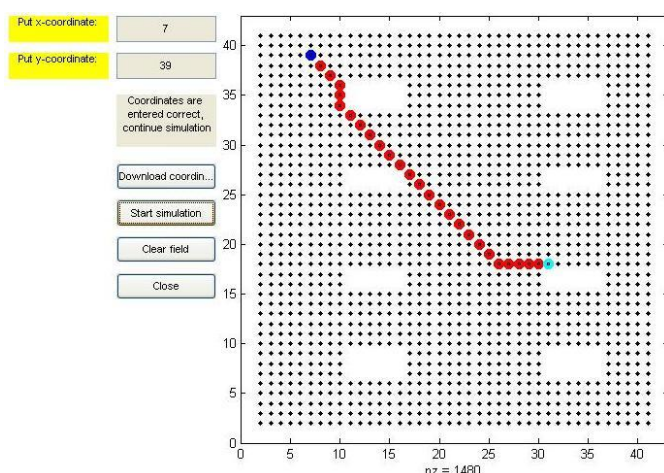


Fig. 5 Other sample of simulation path from start to goal position (start position has coordinates [x=7, y=39])

V. CONCLUSION

Method harmonic potential field is one of the most used methods for path planning of mobile robot in space with obstacles. Robot can find safe path from start to goal positions. Achieved simulation results in this paper indicate that this method is really applicable for path planning of robot. Note simulation results were attained for static space where position of obstacles is fixed. Subject of another study can be to find algorithm for creating harmonic potential field where there is only one global minimum or to extend created program in MATLAB to three-dimensional simulation virtual space.

ACKNOWLEDGMENT



This paper was developed with support of the Project “Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky” (Centre of Excellence of Integrated Research and Exploitation the Advanced Materials and Technologies in the Automotive Electronics), ITMS 26220120055, that is co-financed from Structural Funds EU ERDF within Operational Program Research and Development OPVaV-2009/2.1/03-SORO and preferred axis 2 Support of Research and Development (70%).

REFERENCES

- [1] J. R. Blakely, Potential Theory in Gravity and Magnetic Applications. Cambridge, United Kingdom: Cambridge University Press, 1996. 441 pp. ISBN 0-521-57547-8.
- [2] I. C. Conolly – J. B. Burns – R. Weiss, Path Planning Using Laplace's Equation. Proceedings of the 1990 IEEE International Conference of Robotics and Automation. 13-18 May 1990, Cincinnati, USA. 2102-2106 pp. ISBN 0-8186-9061-5.
- [3] Ch. Shi – M. Zhang – J. Peng, Harmonic Potential Field Method for Autonomous Ship Navigation. Proceedings of the 7th International Conference on Intelligent Transport Systems Telecommunications. 6-8 June 2007, Sophia Antipolis, France. ISBN 1-4244-1178-5.
- [4] J. Vaščák, Využitie potenciálových polí v navigácii mobilných robotov, Technical University in Košice, 2008
- [5] J. C. Latombe, Robot Motion Planning. Massachusetts, USA, Kluwer Academic Publisher, 1991. 651 pp. ISBN 0-7923-9129-2.
- [6] J. Kvasnica, Teorie elektromagnetického pole, Praha, Czech Republik: Academia Praha, 1985. 449 pp.
- [7] K. Y. Hwang – N. Ahuja, A Potential Field Approach to Path Planning. IEEE Transactions on Robotics and Automation vol. 8 no. 1, February 1992
- [8] M. Shahab, Harmonic Potential Field Motion Planning, King Fahd of Petroleum and Minerals, 7 February 2009

Improved experimental model of bi-directional DC/DC Converter

¹Tomáš BÉREŠ (3rd year), ²Michal KALAVSKÝ (3rd year)
Supervisor: ³Jaroslav DUDRIK

^{1,2,3}Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

¹tomas.beres@tuke.sk, ²michal.kalavsky@tuke.sk, ³jaroslav.dudrik@tuke.sk

Abstract— Bi-directional buck-boost DC/DC converter for hybrid battery in electric hybrid vehicle (EHV) controlled by DSC TMS320C28335 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 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), Digital signal controller (DSC)

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 ultracapacitor (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.

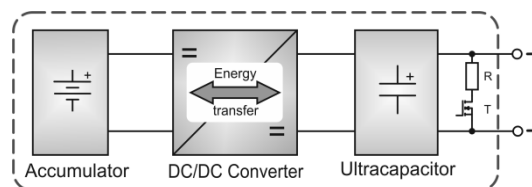


Fig.1 Block diagram of hybrid battery

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 Tab 2.

Parameter	Value
Input voltage U1	10-28 V
Max. output voltage U2	28 V
Max. output voltage ripple ΔUout	1V
Max. output current	30A
Max. current ripple	3A
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.

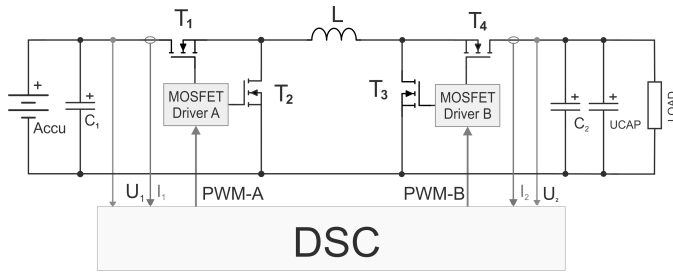


Fig. 2 Topology of bi-directional buck-boost DC/DC converter

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:

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

where:

$$D_1 = \frac{t_{T1(ON)}}{T}; \quad D_2 = \frac{t_{T4(ON)}}{T} \quad (2)$$

$t_{T1(ON)}$ and $t_{T4(ON)}$ indicate the ON time of the MOSFET switches T_1 and T_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 T_4 is always ON and T_3 is always OFF during this mode ($D_2 = 1$). Only T_1 and T_2 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:

$$V_{OUT} = V_{IN} \cdot D_1 \quad (3)$$

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

In this switching mode all four MOSFET's operate during the period. The first path (T_1, T_3 are ON) enables charging the inductor, the second path (T_2, T_4 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 :

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

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

$$V_{OUT} = V_{IN} \cdot \frac{D_1}{1 - D_1} \quad (5)$$

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

Transistor T_1 is always ON and T_2 is always OFF during the period in this mode ($D_1 = 1$). Only T_3 and T_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:

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

B. Control circuit of converter

Control structure with triangle generator, error amplifier and two comparators is used for control of this converter. Scheme of control structure is shown in Fig.3.

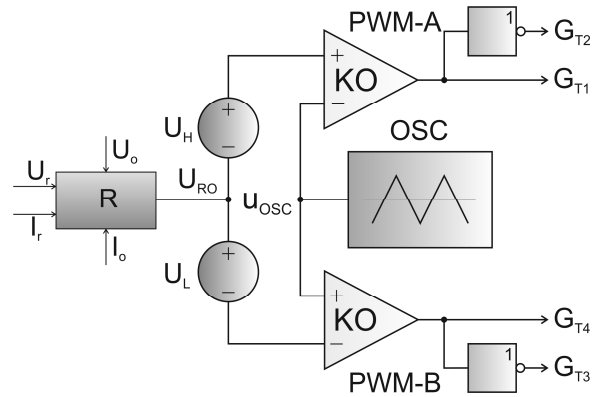


Fig. 3 Control circuit of DC/DC converter

Changing modes of converter depend on value U_H and U_L as follows:

If: $U_{RO} - U_L < U_{OSC_MIN}$ and $U_{RO} + U_H < U_{OSC_MAX}$
buck mode is set

If: $U_{RO} - U_L > U_{OSC_MIN}$ and $U_{RO} + U_H < U_{OSC_MAX}$
buck-boost mode is set

If: $U_{RO} - U_L > U_{OSC_MIN}$ and $U_{RO} + U_H > U_{OSC_MAX}$
boost mode is set

where U_{RO} is output from regulator.

For better understanding the function of control structure is shown in Figure 4.

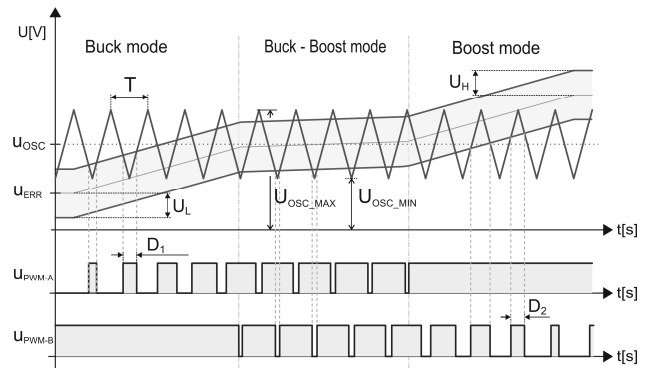


Fig. 4 Function of control circuit

Voltage u_{PWM-A} is input for transistors T_1 and T_2 driver, voltage u_{PWM-B} is input for transistors T_3 and T_4 driver.

Transistors T_1 and T_4 are switched by non-inverted signal and transistors T_2 and T_3 are switched by inverted signal.

IV. EXPERIMENTAL RESULTS

The function of the proposed DC/DC converter was verified on the laboratory model. Whole construction of power modul is mounted on the aluminium PCB, which is attached on the heatsink.

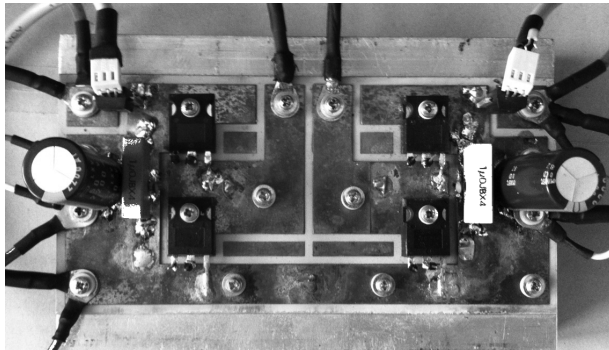


Fig. 5 Experimental model of DC/DC converter

The correct function of bi-directional DC/DC converter was verified with output voltage value 12 Volts. Input voltage U_1 was 24V. Max. current of converter was limited to 30 Amps. Waveforms of input current I_1 , output current I_2 , load current I_Z and waveform of output voltage of DC/DC converter are shown in next oscillograms.

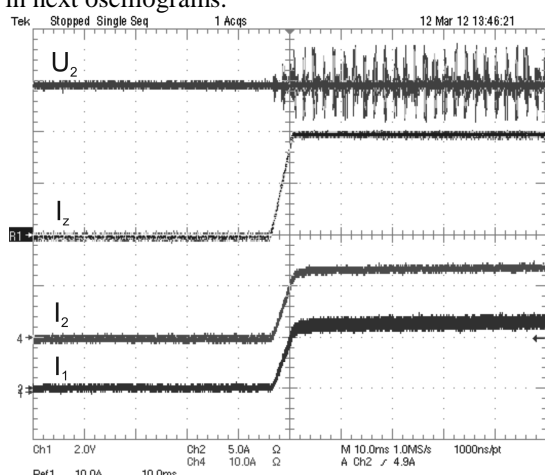


Fig. 6 Waveforms of input current I_1 , output current I_2 , load current I_Z and waveform of output voltage without limitation of the current

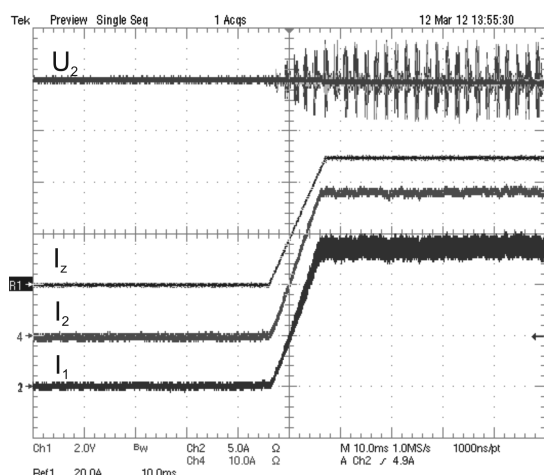


Fig. 7 Waveforms of input current I_1 , output current I_2 , load current I_Z and waveform of output voltage with limitation of the current

Fig. 6 shows function of converter, during step change of load current from zero to 20 Amps. Current I_2 is smaller than maximum current of converter. Limiting of output current I_2 is shown in Fig. 7. In this case is load current 50A.

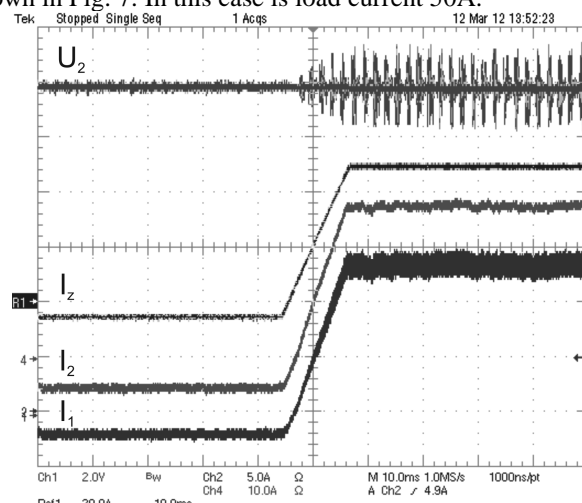


Fig. 8 Waveforms of input current I_1 , output current I_2 , load current I_Z and waveform of output voltage U_2

Bi-directional power flow between accumulator and ultracapacitor is shown in Fig.8. Load current is changed from -5 Amps to 50 Amps.

V. CONCLUSION

Improved model of bi-directional DC/DC converter is described in the paper. The correct function of converter was verified on experimental model and measurement results were shown in last part of this paper.

ACKNOWLEDGMENT

This work was supported by Slovak Research and Development Agency under project APVV-0185-10.

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] Leuchter, J.; Bauer, P.; Bojda, P.; Rerucha, V.: Bi-directional DC-DC converters for supercapacitor based energy buffer for electrical gen-sets. In: Power Electronics and Applications, 2007 European Conference on , vol., no., pp.1-10, 2-5 Sept. 2007.
- [7] Schupbach, R.M.; Balda, J.C.: Comparing DC-DC converters for power management in hybrid electric vehicles. In: Electric Machines and Drives Conference, 2003. IEMDC'03. IEEE International , vol.3, no., pp. 1369- 1374 vol.3, 1-4 June 2003
- [8] Olejár, M.: Hybridná batéria ako sekundárny napájací zdroj s vylepšenou dynamikou výstupného výkonu.(dizertačná práca), TUKE, Košice, 146 s.

Improving transient stability with utilizing FACTS device

¹Zsolt ČONKA, ²Stanislav KUŠNÍR
Supervisor: ³Michal KOLCUN

^{1, 2, 3}Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

¹zsolt.conka@student.tuke.sk, ²stanislav.kusnir@tuke.sk, ³michal.kolcun@tuke.sk

Abstract - This paper describes the impact of TCSC (thyristors controlled series compensator) on the transient stability of the electricity transmission. Line reactance and power flow can be changed by TCSC. TCSC can help to better course of the transient phenomenon ongoing in the power system (PS). TCSC is one of the best ways to improve the transient stability of the system.

Keywords — power system, transient stability, TCSC.

I. INTRODUCTION

Research of the FACTS devices began at the 80thies of the 20th century [1]. One of the main tasks of the regulators was to improve the stability of electricity transmission. Improving the dynamic stability means the increase the maximum transmission capacity in the system without losing sync. [2] The probability of loss of stability depends on the size of the transmitted power before failure. The dynamic stability of power system is influenced by many factors such as: admittance, duration of failure, the regulation of excitation. TCSC allows us to continuously change the admittance, which is a very important variable affecting the stability of the system [3].

II. COUPLING ADMITTANCE

The coupling admittance Y_{ij} indicate as strongly are the generators electrically linked each other. Enlargement of the admittance, before, after or during disturbances has a positive effect on the transient stability of the system. If we want to increase the coupling admittance we must reduce the coupling impedance. The coupling impedance Z_{ij} is determined by a combination of individual elements of the PS. It follows that to improve the stability it is necessary to reduce the reactance of the generators, transformers and lines. Reducing the reactance of the existing machines are impossible. Reducing of the line reactance is possible with using a bunch lines. Using of the bunch lines, the reactance of lines are reduced by 20% to 30%. Another possibility of indirect impedance changes is the inclusion of different compensators and FACTS devices [2].

III. TCSC

TCSC consists of a capacitor which is parallel connected to a thyristor regulated reactor. This configuration allows you to smoothly change the impedance of the transmission lines and smooth control of frequency of the capacitance in a wide range. Control is very important for proper function of TCSC.

The control of the reactance of TCSC is very complex and dynamic process. Efficiency and effect of TCSC are very dependent on the accuracy of calculation of the TCSC reactance at the moment. In case of wrong calculation of the reactance, may also occur to the opposite effect than what is provided.

TCSC consists of 3 parts: capacitor battery C, parallel connected reactor L and the thyristors VT1 and VT2 (Fig. 1). [5]

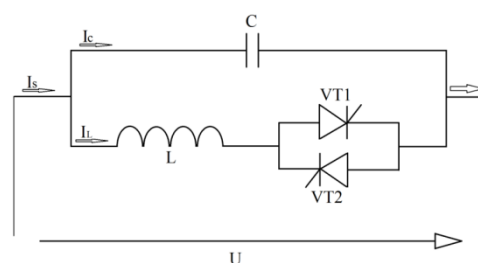


Fig. 1 Engagement of TCSC

$$i_c = C \cdot \frac{dv}{dt} \quad (1)$$

$$i_L = L \cdot \frac{di_L}{dt} \quad (2)$$

$$i_s = i_c + i_L \quad (3)$$

where:

I_C and I_L - Is the instantaneous capacitor and reactor current.

i_s - Is the instantaneous current of controlled line.

U - Is the voltage of TCSC.

TCSC can be controlled by changing the opening angle of the thyristors α , which is modifying the frequency of the

capacitor. Relationship (1) gives the relationship between the opening angle of the thyristors (α) and reactance $X_{TCSC}(\alpha)$.

$$X_{TCSC}(\alpha) = X_C - \frac{X_C^2}{(X_C - X_P)} \cdot \frac{\sigma + \sin \sigma}{\pi} + \frac{4X_C^2}{(X_C - X_P)} \cdot \frac{\cos^2\left(\frac{\sigma}{2}\right)}{(k^2 - 1)} \cdot \frac{\left(k \cdot \tan\left(\frac{k \cdot \sigma}{2}\right) - \tan\left(\frac{\sigma}{2}\right)\right)}{\pi}, \quad (4)$$

where:

X_C - Is the capacitor capacitance.

X_P - Is the inductive reactance of the reactor connected in parallel to the condenser.

$\sigma = 2(\pi - \alpha)$ = opening angle of TCSC controller

$k = \sqrt{\frac{X_C}{X_P}}$ = compensation ratio.

$X_{TCSC}(\alpha)$ is a unique feature by which we can continuously change the TCSC reactance. TCSC can operate in capacitive or inductive mode, but the transition from one mode to another we must avoid areas of resonance. TCSC is modeled as a variable capacitor, which is limited only by the regulatory scope of the angle α .

The Control range of TCSC is:

$$X_{TCSC}(\min) \leq X_{TCSC}(\alpha) \leq X_{TCSC}(\max), \quad (5)$$

where:

$X_{TCSC}(\min) = X_{TCSC}(180^\circ)$ - Thyristor in permeable state.

$X_{TCSC}(\max) = X_{TCSC}(\alpha_{\min})$ - Thyristor in trailing state [5, 6].

IV. TRANSIENT STABILITY

Synchronous machines are mainly used in the PS as synchronous generators, synchronous motors and synchronous compensators. These devices are interconnected through transformers and lines and they are in parallel and synchronous operation. The transfer capability of such systems is limited by the permissible voltage drops and with the power handling capacity. At transmission to the large distance been treated the condition of stability of parallel operation. Prerequisite for the existence and operation of large interconnected power systems is the steady stability of parallel operation that is to say sync. Synchronization power of synchronous machine allows us to so. synchronous operation. This power presents the increase of the transmitted power at an increasing of the load angle of generator rotor over 1° [2, 3]. The swing of generator rotor due to changes of the electromagnetic energy which is accumulated in the magnetic circuits of the machine so arises additional power. This power can take positive but also negative values, which is reflected as a contribution to the performance of ΔP as a braking, or acceleration of the machine.

The role of synchronous machines not the retention of the synchronism. Co-operation of synchronous machines cannot be at arbitrarily large power. The Steady-state operation of whole system is depends on the electrical, mechanical and electromagnetic parameters of the system.

If you are not violated conditions of transfer we can reach a stable maximum output with the smoothly increasing transmitted power. The considerations of static stability we can apply for small swings of the machine. In operation, there are many sudden changes in the system (switching processes, shock loads, short circuits). Due to the unbalance of the consumption and production of the electric power the load angle values can achieved a major changes. These marches change the load angle with leap. After these changes the system goes into the new system state with electromechanical oscillations. Inertia of the machine not allow an immediate change of operating parameters (angle δ) course of these oscillations may be such as that the angle stabilizes at a new constant, or continue to increasing. In the latter case there is a loss of stability. In examining of the transient stability is generally assumed a constant value of the transition electromotive alternator voltage E [2, 3].

V. EQUAL AREA CRITERION

The principle of establishing transient stability using the method of a surface is explained on a simple network. Synchronous generator is supplied electricity to the grid. For explanation we introduce an example of failure of one of the parallel lines. The power of the generator is transferred to the grid by only one line.

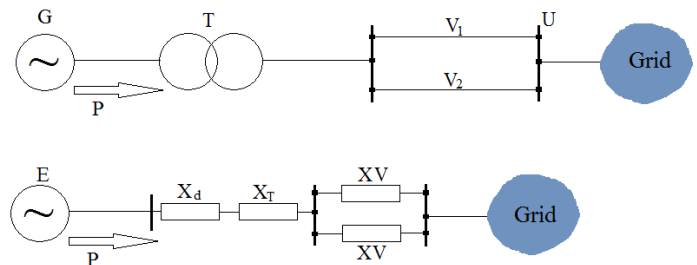


Fig.2 Engagement of network

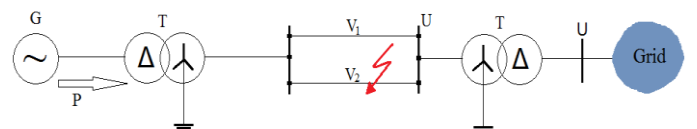


Fig. 3 Fault of the line

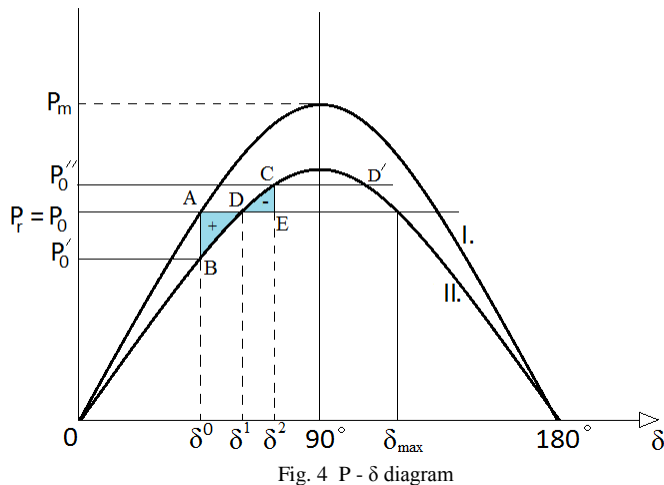


Fig. 4 P - δ diagram

The new operating condition is characterized in that there is a failure of one of the parallel lines leads to the increase in reactance from X_0 to $X_1 = 2 \cdot X_0$. Angle ϑ_0 and mechanical power of aggregate P_m are not changed due to inertia of the rotating masses of the generator. This fact is shows the decrease of the electrical power is appears as the acceleration torque i.e. accelerator power

$$P_a = P_m - (P_0 - \Delta P) \quad (6)$$

This change is reflected as the decline of power from the point A(P_0, ϑ_0) to point B($P_0 - \Delta P, \vartheta_0$) on power characteristics at the same angle ϑ_0 . The acceleration and the acceleration power of the machine brings operation to the point D(P_0, ϑ_1). Point D on the power characteristics shows the operation status with one line with reactance $X_1 = 2 \cdot X_0$. This transfer would be possible at an angle ϑ_1 . The difference between the angle ϑ_1 and ϑ_k would be considerably smaller than the ϑ_0 in this case [2].

In the Point D the generator have greater angular velocity than synchronous angular velocity. The status is still unstable. Power is greater than the load and the load angle increases further to state of corresponding to point C ($P_0 + \Delta P, \vartheta_2$). At this point, the mechanical power of the machine is smaller than a load. The generator is slowing here, until the point C where enters into synchronous rotation. This example illustrates the fact that a rapid change of load near the critical angle of load can induce a state, where the return to sync. is not possible [2].

VI. SIMULATION

On simple example with synchronous generator we show the impact of TCSC on the transient stability. The power is transferred to system by parallel lines. At time $t = 0.2s$ will due to short-circuit failure of one of the parallel lines. Due to the failure comes to a rotor swings. This example is simulated in the simulation program NEPLAN. In the first case, the simulation is without TCSC and in the second case involved the TCSC.

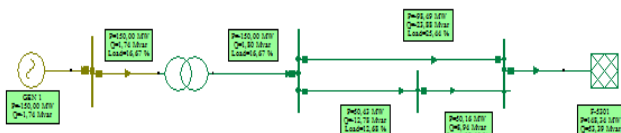


Fig. 5 Engagement without TCSC

Automatic control is used to control of TCSC (Fig.6). The actual transmitted power with lines is the input to the controller. Limiter is another part of the regulator. Limiter has set a maximum possible value of the transmitted power with one line. Opening angle of the thyristors is controlled by a regulator. The controller managed the reactance of TCSC according to the current transmitted power.

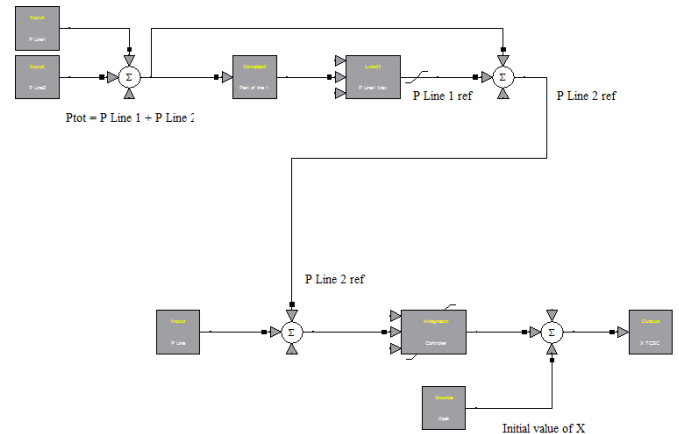


Fig.6 TCSC controller

Failure of one of the parallel lines occurs due to short circuit. On figure 7 we see that the power generator begins to swing and fall out of sync.

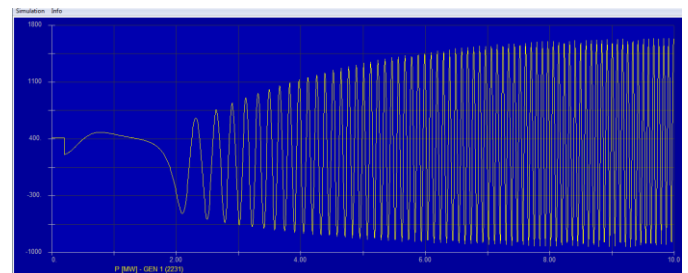


Fig. 7 Power swing without TCSC

On Figure 8 we see the networks with the TCSC. The power of the generator after fault begins to swing. TCSC will prevent the loss of sync of the generator. Swing of the generator is stabilized after a few seconds and the network operates with one line.

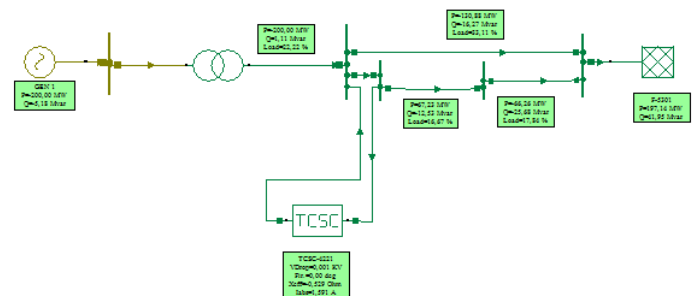


Fig. 8 Engagement with TCSC

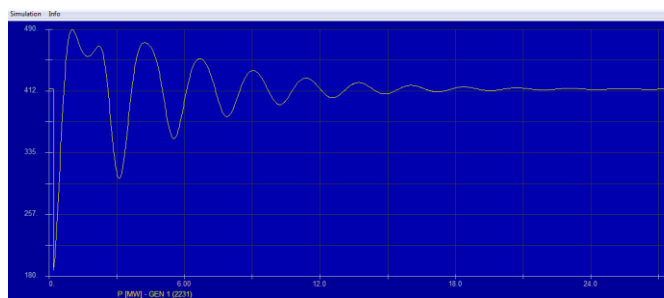


Fig. 9 Power swing with TCSC

TCSC has a positive effect on the transient stability of the system and can help to get the better reliability of electricity supply. If we prevent loss of sync, we can prevent "chain reaction". This means that a failure of the single generator with big power may lead to further overload of the next generator and to his subsequent failure to. This "chain reaction" can lead to worst possible situation and that is Black out.

References

- [1] M. A. ABIDO. Power system stability enhacement using FACTS controllers [online]. [cit. 2012-02-05]. Available on the Internet: <http://ajse.kfupm.edu.sa/articles/341B_P.12.pdf>
- [2] TROJÁNEK, Zdeněk – HÁJEK, Josef – KVASNICA, Pavol: Přechodné jevy v elektrizačních soustavách. STNL , 1987. 202 – 231 s.
- [3] REISS, L., MALÝ, K., PAVLÍČEK, Z., BÍZIK, J.: Teoretická elektroenergetika II. STNL Praha, Alfa Bratislava 1978
- [4] CINTULA B. Analýza dynamickej stability synchronného generátora [online]. [cit. 2012-02-12]. Available on the Internet: <<http://www.posterus.sk/?p=11726>>.
- [5] FACTS – Powerful systems for flexible power transmission. [online]. <[http://www05.abb.com/global/scot/scot221.nsf/veritydisplay/b0f2c8c94b48a6bcc1256fda003b4d42/\\$file/facts_%20eng.%20abb%20review.pdf](http://www05.abb.com/global/scot/scot221.nsf/veritydisplay/b0f2c8c94b48a6bcc1256fda003b4d42/$file/facts_%20eng.%20abb%20review.pdf)>
- [6] NARAIN, G. HINGORANI, LASZLO, GYUGYI.: Understanding FACTS. ISBN: 0-7803-3455-8, IEEE Order No. PC5713
- [7] REVÁKOVA, D., ELESCHOVÁ, Ž., BELÁŇ, A.: Prechodné javy v elektrizačných sústavách. ISBN, 978-80-227-2868-3

Influence of PST transformers on short-circuit condition of the electric power system

¹Vladimír KRIŠTOF (3th year), ²Stanislav KUŠNÍR (3th year),
Supervisor : ³Michal KOLCUN

¹Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
^{2,3}Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

¹vladimir.kristof@tuke.sk, ²stanislav.kusnir@tuke.sk, ³michal.kolcun@tuke.sk

Abstract—Nowadays remains a growing trend of energy consumption. 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. This problem affects a lot of countries, including Slovakia. Since building new lines is very tedious process, one of the solution is using special devices designed to control power flow, such as PST, UPFC or FACTS. Introduction of new elements used to regulate the load flow (eg PST transformers, FACTS devices, HVDC) also affects the value of short circuit currents. Short-circuit condition is one of the most important aspects which must be taken into account in a development of power system. This article deals mainly with change of short-circuit conditions in case of use of PST transformers in the transmission system

Keywords—PST transformers, Short-circuit condition, Power system

I. INTRODUCTION

Power system of the Slovak Republic is a part of the European interconnected systems ENTSO-E (European network of transmission system operators for electricity) . Attention is currently devoted especially to safe and reliable operation and control of power system operated under ENTSO-E. Increasing trend in energy consumption causes higher use of transmission elements. A look at investment opportunity of development, there are used new elements (PST transformers, HVDC etc.) to control power flows in operated transmission system . These elements except good control properties, also influences other areas such as short-circuit conditions, loss or transient stability of power system.

The term short-circuit means undesirable conductive connection between phases, phase and ground, which leads to a reduction of impedance in electrical circuits, and thus the flow of undesirable short-circuit currents. It is very important to know maximum value of short-circuit current (for dimensioning of electrical equipment) and minimum value of short-circuit current (for setting-up of protection devices like distance protection) of short-circuit current[2]. Knowledge of short-circuit conditions in the power system is a necessary for its safe operation.

PST transformers (phase shifting transformers) are used in

transmission systems to control flows in power lines. Voltage regulation is only additional feature. Modern thyristor controlled PST can also be used to improve the transient stability or power oscillations in power system. Transformers, are usually placed into one voltage level, often in outlet lines. Phasor angles change the output voltage to input voltage on the angle switch branches [3,5,6]. Phasor diagram of PST transformer is shown on fig.1.

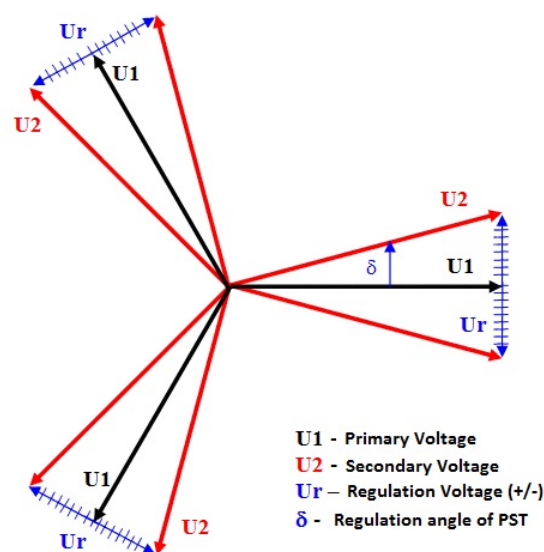


Fig. 1. Phasor diagram of PST Transformer [3]

This article analyzes an impact of PST transformer on short-circuit conditions in the power system.

II. USE OF PST TRANSFORMERS IN POWER SYSTEM OF SLOVAK REPUBLIC

Transmission system of the Slovak Republic is interconnected with foreign systems via 400 and 220 kV power lines. Taken into the account the fact that the 220 kV level has no future in the Slovak transmission system, in this article is

considered only 400 kV level. At this level are operated the following overhead lines:

- V477, V478: Lemešany – Krosno (Poland) (double 400 kV overhead line)
- V440: Veľké Kapušany – Mukačevo (Ukraine)
- V404: Varín – Nošovice (Czech Republic)
- V497: Stupava – Sokolnice (Czech Republic)
- V424: Křižovany – Sokolnice (Czech Republic)
- V448: Gabčíkovo – Győr (Hungary)
- V449: Levice – Göd (Hungary)

In this article attention is paid especially to most loaded power lines with Czech Republic and Hungary.

TABLE I
SHORT-CIRCUIT CONDITIONS IN BORDERS SUBSTATIONS WITHOUT PST TRANSFORMER

Substation	$I''_{k3}(\text{kA})$
Varín	13,05
Stupava	10,24
Křižovany	14,76
Gabčíkovo	10,55
Levice	12,24

PST transformers are put into these border substations to the outlet of power lines with the following parameters:

$S_n = 1\,630\text{ MVA}$

$U_n = 400\text{ kV}$

$R_{p-s} = 0,0019\text{ p.j.}$

$X_{p-s} = 0,110025\text{ p.j.}$

Phase shift: from -40° to $+40^\circ$

Change of the short-circuit conditions in various border substations when PST is considered (Table II and Table III):

TABLE II
SHORT-CIRCUIT CONDITIONS IN BORDERS SUBSTATIONS WITH PST TRANSFORMER

Substation	$I''_{k3}(\text{kA})$				
Phase shift	-40°	-30°	-20°	-10°	0°
Varín	12,13	12,07	12,03	12	11,99
Stupava	9,37	9,35	9,34	9,33	9,32
Křižovany	13,91	13,84	13,79	13,76	13,75
Gabčíkovo	9,69	9,76	9,66	9,65	9,64
Levice	11,58	11,52	11,48	11,46	11,45

TABLE III
SHORT-CIRCUIT CONDITIONS IN BORDERS SUBSTATIONS WITH PST TRANSFORMER

Substation	$I''_{k3}(\text{kA})$			
Phase shift	10°	20°	30°	40°
Varín	12	12,05	12,09	12,1
Stupava	9,33	9,33	9,34	9,35
Křižovany	13,76	13,79	13,84	13,9
Gabčíkovo	9,65	9,66	9,67	9,69
Levice	11,46	11,48	11,52	11,56

Percentual change of short circuit conditions in the various substations are expressed graphically:

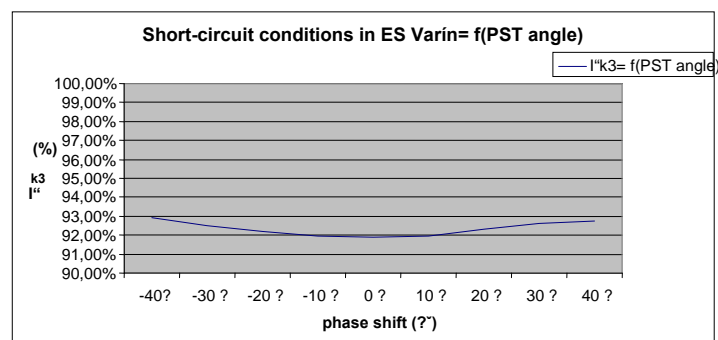


Fig. 2. Change of Short-circuit conditions in substation Varín

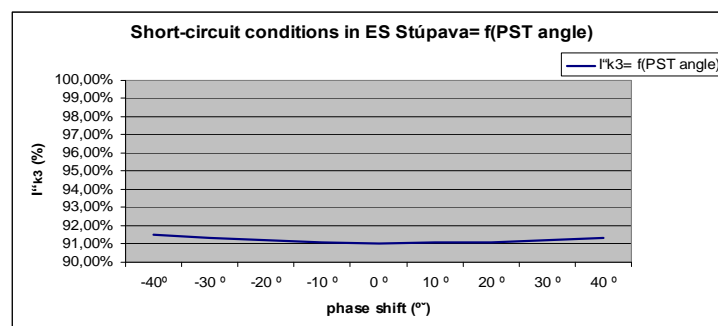


Fig. 3. Change of Short-circuit conditions in substation Stupava

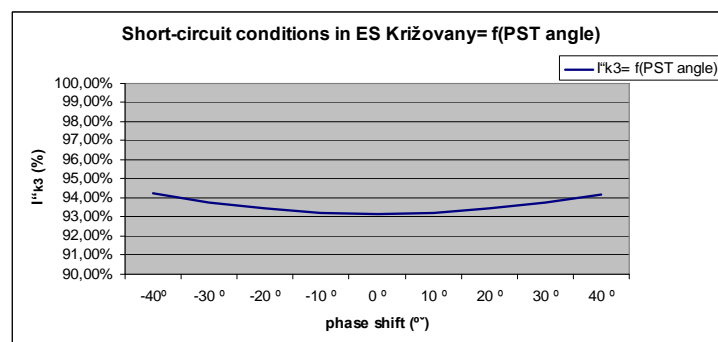


Fig. 4. Change of Short-circuit conditions in substation Křižovany

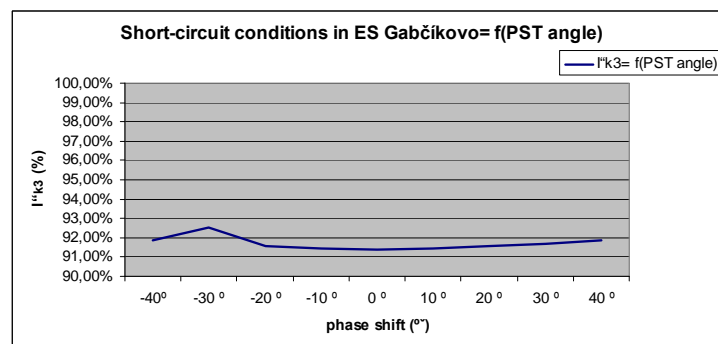


Fig. 5. Change of Short-circuit conditions in substation Gabčíkovo

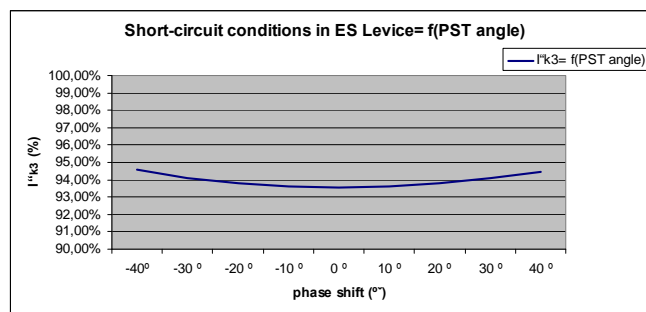


Fig. 6. Change of Short-circuit conditions in substation Levice

Putting PST on these power lines caused not only the possibility of better control of power flows, but also reduced short-circuit currents in substations. Improved short-circuit conditions is significant from the point of view of operational safety but also from the point of view of further development of the transmission system.

III. CONCLUSION

The article deals with the influence of PST transformers on the short-circuit conditions of the transmission system and change of these values. Analyzed were 400 kV border substation operated by SEPS a.s. (Transmission operator of Slovak Power system). In article was shown that PST transformers except good power flow control has influence also on short circuit conditions. It is very important to say, that it is only secondary benefit of PST.

IV. ACKNOWLEDGMENT

This work was supported by Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences under the contract No. 1/0166/10

V. APPENDIX

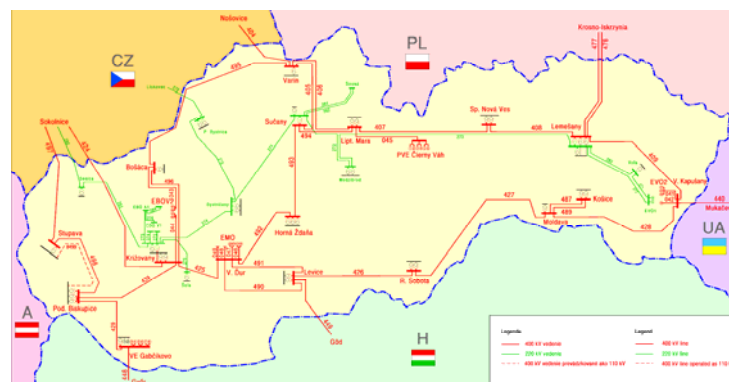


Fig. 7. Power system of Slovak republic

REFERENCES

- [1] Krištof, V.: *Riadenie ES v krízových situáciach*. Minimová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2010.
- [2] Mešter, M.: *Výpočet skratových prúdov v trojfázových striedavých sústavách*. ABB-elektro, s.r.o., 2005. ISBN 80-89057-10-1
- [3] Kušnir, S.: *Možnosti regulácie tokov výkonov v elektrizačných sústavách*. Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2009.
- [4] STN IEC 60909: *Skratové prúdy v trojfázových striedavých sústavách*. Slovenský ústav technickej normalizácie, apríl 2003.
- [5] Beňa, E.: *Využitie špecializovaných zariadení na reguláciu tokov činných výkonov v elektrizačných sústavách*. Habilitačná práca. Košice: Technická univerzita v Košiciach, 2010, 84 s.
- [6] Máslo, K. – Kasembe, A.: *Matematické metódy využiteľné pro dispečerské řízení – vliv transformátoru s regulací fáze*. V mezinárodní vědecké sympozium ELEKTROENERGETIKA 2009, Stará Lesná
- [7] Operator of Slovak transmission system official site: www.sepsas.sk

Influence of the automatic transfer switch (ATS) on the power supply reliability

¹Marián HALAJ (4th year)
Supervisor: ²Ladislav Varga

¹Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

²Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

¹marian.halaj@centrum.sk, ²ladislav.varga@tuke.sk

Abstract—This paper deals with the power supply reliability improvement using automatic transfer switch (ATS). The study investigates the advantages of ATS installation in upstream and downstream side of 33/11kV distribution substation.

Keywords—Distribution substation, reliability, automatic transfer switch, power supply.

I. INTRODUCTION

One of the main criteria that the engineers have to consider during the design of electrical systems is the reliability of power supply. Proper design of distribution substations have to respect the requirements of customer, applicable local and international standards, as well as the economical criteria. This task is a multi-criteria decision making process where the importance of each criteria depends from project to project. In case the consumption doesn't require the highest level of reliability (i.e. aluminium or glass industry) what is the case of utility distribution substation, the owner of the substation has an option to optimize the design from the investment point of view. The task of this optimization process is to find the minimum investment and operation cost while the certain level of power supply reliability to the consumer is kept. Other point of view is to maximize the power supply reliability of the plant.

II. SUBSTATION CONFIGURATION

The subject of this analyze is existing 33/11kV distribution substation (Fig.1). Upstream side is supplied from 2 cable feeders from sub-transmission substation. Two gas insulated primary switchgears are may by connected by means of bus coupler circuit breaker with automatic transfer source ATS1.

Downstream 11kV power distribution is made of two air insulated switchgears that may be connected by means of vacuum circuit breaker with automatic transfer source ATS2.

In normal operation, both upstream and downstream bus coupling circuit breakers are open. Section No. 1 of downstream switchgear is fed from section No. 1 of upstream switchgear through related 33/11,5kV power transformer. Similar situation is applicable for the section No.2. In case of

failure on 33kV incomer the ATS1 will transfer the power supply from second section of the 33kV switchgear. Same scenario will happen in case of 11kV incoming feeder failure. Since the procedure is valid for the power supply failure on second section as well, there are four modes of operation available.

The purpose of the study is to evaluate the reliability of the system and identify the impact of ATS on the reliability of the system.

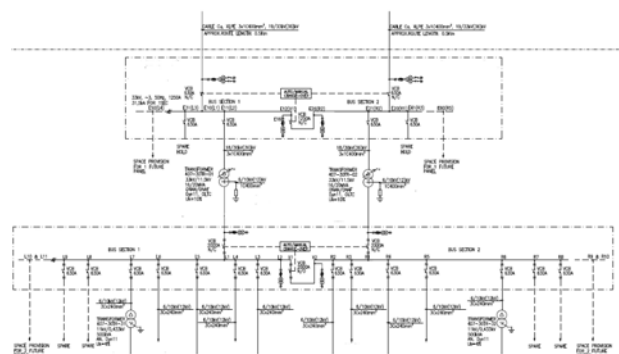


Fig. 1. Single Line Diagram, Substation 33/11kV.

III. AUTOMATIC TRANSFER SWITCH

A transfer switch is an electrical switch that reconnects electric power source from its primary source to a back up or standby source. These switches may be manually or automatically operated. An Automatic Transfer Switch (ATS) is normally installed where the very fast transfer is required or where no operators available.

The transfer switch isolates the failed power supply path from the rest of the system and further connects to the healthy power supply that is normally in standby. To perform this function the ATS is designed as a complex system consisting of synchronization modules, metering, programmable control devices and interface to the circuit breakers. The reliability of ATS depends on the number and quality of the individual components as well as on the system configuration. Principle ATS configuration is shown on the Fig. 1. Bus coupling circuit breaker is normally in open position. In case of failure on Incoming 1 the ATS automatically open I204 breaker and

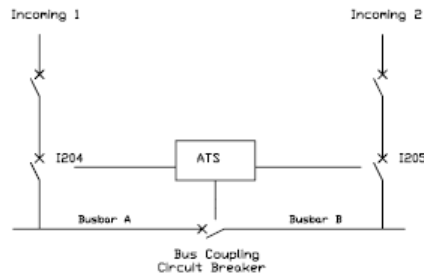


Fig. 2. Typical ATS configuration

closes the coupling and the I205 as well.

IV. RELIABILITY BACKGROUND

A. Reliability Theory

Bayes' theorem method has been chosen to evaluate the reliability of the system. This theory says that the reliability of a system is equal to the reliability of the system, given that a chosen unit is good, multiplied by reliability of this unit, plus the reliability of the system, plus the reliability of the system, given that unit is bad, multiplied the unreliability of this unit [1]. Mathematically we can write this formula as:

$$R_s = R_s \left(\frac{R_s}{A_g} \right) R_a + R_s \left(\frac{R_s}{A_b} \right) Q_a \quad (1)$$

B. Component Reliability Data

The reliability data taken into consideration during the system reliability evaluation have been taken from IEEE 493 Gold Book. These reliability data have been collected from equipment reliability surveys and a data collection program over a period of 35 years or more. This data is the most comprehensive database publicly available on electrical and mechanical equipment reliability in the world [2]. Table 1 shows the reliability indexes of the power components.

Component	λ (failures per year)
MV circuit breaker	0,9826
Power transformer	0,9848
Distribution transformer	0,9941
33kV Cable	0,9965
11kV Cable	0,9939
ATS 1	0,9178
ATS 2	0,9686
Upstream network	0,9103

Table 1. Component reliability data

V. RELIABILITY MODELLING OF THE OPERATION MODES

To evaluate the contribution of each of two ATS three reliability models of the substation has been performed:

- Parallel redundant system without ATS
- System with upstream ATS1 active
- System with downstream ATS2 active
- System with both ATS1 & ATS2 active

	Net	C1	C2	VCB1	VCB2	ATS1	VCB3	C3	VCB4	C4	TR01	TR02	C5	VCB5	C6	VCB6	ATS2	VCB7	C7
Case 1	1	1	0	1	0	0	1	1	0	0	1	0	1	1	0	0	0	1	1
Case 2	1	0	1	0	1	0	0	0	1	1	0	1	0	0	0	1	1	1	1
Case 3	1	1	0	1	0	1	0	0	1	1	0	1	0	0	1	1	1	1	1
Case 4	1	0	1	0	1	1	1	1	0	0	1	0	1	1	0	0	0	1	1

Table 2. Power supply to feeder L7 - operation procedure

C1-C7 Cables

VCB1-VCB7 Circuit breakers

TR01-TR31 Transformers

ATS1 Upstream automatic transfer switch

ATS2 Downstream automatic transfer switch

The operation procedure of each case of operation shows Table 2. Digit 1 means the equipment is in operation, 0 means it is not in operation/no loaded. In first case there is no ATS considered. Distribution system is made of two independent power supplies. Reliability is calculated on the secondary terminal of the distribution transformer TR31 (reliability of the substation auxiliary supply), therefore the parallel redundancy of the system doesn't have any impact on the reliability value. The explanation is that we have to consider one transformer out of service due to maintenance or other reason. Each of them has one 33kV incoming feeder and one 33/11,5 kV 20MVA power transformer with on-load tap changer. Impact of tap control and transformer protection is included in transformer reliability index value and it is not considered in the model separately.

VI. RESULTS OF THE CALCULATION

Table 3 provides the calculated reliability of different substation configuration. Simple radial system without any automatic transfer has significantly lower level of the reliability than the other configurations. Using one automatic transfer switch either on downstream or on upstream side highly improves the total level of reliability. Reliability of the configuration 2 (case 2) has a bit higher level of the reliability. This result may be explained with the fact that in this configuration there are more back up components comparing to configuration case 2. Application of second automatic transfer switch doesn't have high influence on the further improvement of the reliability. This result may be likely explained with the fact that the automatic transfer switches as a components of the reliability scheme have much higher failure rate (lower reliability of the components) than other components. For example the reliability of ATS1 is only 0.9178 comparing to the reliability of distribution transformer that is 0.9941. Generally, with the increasing of the numbers of components in the configuration the reliability decreases. This is, specially, valid for the electronic components like automatic transfer or control system.

Configuration	Reliability
Case 1 (Simple radial system)	0,81318
Case 2 (Downstream ATS2)	0,89722
Case 3 (Upstream ATS1)	0,89281
Case 4 (ATS1 & ATS2)	0,90423

Table 3. Calculated reliability of different substation configuration

Another problem is the input reliability of the substation. Since 80% of all the power interruptions is arisen in distribution networks, we can't consider the substation as isolated system. We consider the reliability of the upstream network of 0.9103 what is even worst value than the reliability of ATS.

Configuration	λ (failure per year)
Case 1 (Simple radial system)	0,207
Case 2 (Downstream ATS2)	0,108
Case 3 (Upstream ATS1)	0,113
Case 4 (ATS1 & ATS2)	0,101

Table 4. failure rate of different substation configuration

Configuration	Annual cost of interruption
Case 1 (Simple radial system)	20.700
Case 2 (Downstream ATS2)	10800
Case 3 (Upstream ATS1)	11300
Case 4 (ATS1 & ATS2)	10100

Table 5. Annual cost of interruption in EUR

From the exponential distribution of the failures can be calculated the failure rate (Table 4). This is one of the most important information for the reliable substation design with the economical constrain. In case the commercial information about the penalties for not supplied energy available the design engineers can compare the prices with the cost of ATS. Then he can decide, if or how many ATS switches is effective to install. Similar approach may be used to evaluate the substation down time per year. In our case we consider 100.000EUR per interruption due to lost of production. In this case the total cost of interruption per year is 20.700EUR in case of simple radial system, 11.300EUR with upstream ATS only and 10.100EUR in case of 2 ATS modules installed.

VII. CONCLUSION

The substation configuration has critical impact on the number of failures and substation down time per year. Since there is a direct proportionality between the number of failures and cost of unsupplied energy that have to pay the owner of the substation to the customers based on the valid contract, described procedure gives owner the guidance to optimize these costs. In this case the annually saved with 2 ATS switches is 10.600 EUR against the two simple radial systems, but only 700EUR saved against one ATS, therefore the installation of 2 ATS switches is not cost effective.

REFERENCES

- [1] D. Kececioglu, Reliability Engineering, Handbook. Lancaster: DEStech Publications, 2002, p. 535, ISBN 1-932078-01-0.
- [2] IEEE 493: 2007, IEEE Recommended Practice for Design of Reliable Industrial and Commercial Power Systems.
- [3] IEC 61078: 2006, Analysis techniques for dependability – Reliability block diagram and boolean methods.
- [4] The IEEE Standard Dictionary of Electrical and Electronics Terms, IEEE Std.100-1996
- [5] J. Tuma, S. Rusek, Z. Martinek, M. Tesařová, I. Chmišinec, Security, quality and reliability of electrical energy, Conte spol. s.r.o., ČVUT Praha, 2007, 172 s., ISBN 978-80-239-9056-0.

Interpretation of regular force measurements in fluids

¹Gabriela VASZIOVÁ (3rd year), ²Lukáš GLOD (4th year)
Supervisor: ¹Vladimír LISÝ

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, ¹vladimir.lisy@tuke.sk

Abstract— In this paper we explore the problem of the determination of regular forces acting on microscopic and smaller objects in fluids. Regular forces can be determined, for example by the measurement of the drift velocity of Brownian particles. In this contribution we have obtained an exact expression for this velocity within the hydrodynamic theory of the Brownian motion. It is shown that the influence of the inertial and memory effects can be significant in the force determination when the experimental times are sufficiently short.

Keywords— Langevin equation, regular forces, Brownian motion.

I. INTRODUCTION

Nowadays, the interest to the investigations of systems at small space and time scales is increasing; this is the reason why the experimental methods of the determination of ultra-small forces (down to femtonewtons) attract great attention. At present, new methods of such measurements are being developed [1]. The principal problem in these studies on micro- and nano-objects is connected with the influence of omnipresent thermal noise. As distinct from macroscopic systems, the noise can essentially affect the motion of small objects and the forces originating from the noise can even exceed the conservative forces in the system. This can lead to a completely incorrect interpretation of the experiments. This question is not solved up today and there is an interesting discussion about it in the literature [1 - 4]. Two methods of force measurements have been used in [1, 4] to determine the influence of thermal noise on a colloidal particle near a wall in the presence of electrostatic and gravitational fields. One of the methods consisted in measuring the particle drift velocity and the subsequent use of the overdamped Langevin equation [5] to determine the full force acting on the particle. In the second method the force was determined from the known particle-wall interaction potential using the equilibrium Boltzmann distribution. The results of these methods were strongly contrasting: the obtained forces deviated both in their magnitudes and even in their sign. As a resolution of these discrepancies, criticized however in [2], the choice of the anti-Itô conception of stochastic calculus has been suggested for the case of spatially inhomogeneous diffusion. In the present contribution we address one more problem of the measurements of forces acting on microscopic and nanoscopic objects that exists even in the case of homogeneous diffusion

(for unbounded particles and constant forces). This problem arises with lowering the time of measurement when the particles undergo unsteady motion for which the standard Langevin description becomes inappropriate. We show how due to the hydrodynamic aftereffect the experimentally determined force, even if it is known to be constant, appears to depend on the relation between the time of measurement and a characteristic time of the loss of memory in the system.

II. ON THE REGULAR FORCE MEASUREMENTS AT SHORT TIME SCALES

It follows from the overdamped Langevin equation that when a microscopic body is suspended in a liquid, a force F applied to a body results in a drift velocity $v = F/\gamma$, where γ is the object's friction coefficient. This is true for large constant forces, if the inertial effects are neglected [1, 4]. However, when the drift force amplitude is comparable to the effect of thermal noise and the force depends on the particle position, the equation for F must be corrected [1] since the measured velocities are statistically distributed. Moreover, an additional term $-\alpha\gamma(x)dD(x)/dx$, referred to as *spurious force*, should be added to the force $F(x) = \gamma(x)\langle v(x) \rangle$ (for the force changing in the direction x). Here, $D(x)$ is the position-dependent diffusion coefficient and α is a constant from the interval $[0, 1]$. For Brownian particles the preferred value is $\alpha = 1$ [5] but there is no common agreement as to this choice [2, 6, 7], which significantly affects the stochastic calculus. One more problem, to our knowledge not considered so far in the interpretations of force measurements, appears in situations when inertial effects (and, consequently, the memory in the particle dynamics) can play a role [8]. Even if the applied force is position independent, but the observation times become comparable to the characteristic time of the loss of memory in the system (the vorticity time $\tau_R = R^2\rho/\eta$, where R is the particle radius and ρ and η are the density and viscosity of the solvent), the discussed method is not applicable. We will show it coming from the generalized Langevin equation [5, 8, 9],

$$M\dot{v}_i(t) + \gamma v_i(t) + \int_0^t \Gamma(t-t')\dot{v}_i(t')dt' = F + \zeta(t). \quad (1)$$

Here, $\zeta(t)$ is a random noise force with zero mean driving the

particles of mass M_p ($M = M_p + M_s/2$, M_s is the mass of the solvent displaced by the particle), and F is a regular force to be determined. The force ζ is, due to the fluctuation dissipation theorem (FDT), connected to the dissipative properties of the system. We consider the very realistic case when the memory in the system is of the hydrodynamic kind, i.e., the resistance force against the particle motion (the Boussinesq-Basset force, here rewritten as the integral from zero to t) follows from the non-stationary Navier-Stokes equations of motion for incompressible fluids [10, 11]. Then the kernel Γ is $\Gamma(t) = \gamma (\tau_R/\pi t)^{1/2}$. The usual Brownian relaxation time is connected to the Stokes friction coefficient as $\tau = M/\gamma$. Equation (1) for $F = 0$ describes the zero-mean fluctuations $v(t)$. Here we are interested in the question how the thermal noise influences the determination of the force F . Let F be constant, as it is for a freely falling particle in a fluid. We express the velocity as $v_t = v + v^*$. The deterministic part v^* (the drift velocity) obeys the averaged equation (1), i.e. the equation without the random force

$$\dot{v}^* + \frac{1}{\tau} v^* + \frac{1}{\tau} \sqrt{\frac{\tau_R}{\pi}} \int_0^t \frac{\dot{v}^*(t')}{\sqrt{t-t'}} dt' = \frac{F}{M} \quad (2)$$

The initial condition at $t = 0$ (when the force F begins to act) is $v^*(0) = 0$. Then the Laplace-transformed (2) has the following solution for $\tilde{v}^*(s) = \mathcal{L}\{v^*(t)\}$:

$$\tilde{v}^*(s) = \frac{F}{Ms} \frac{1}{\lambda_1 - \lambda_2} \left(\frac{1}{\sqrt{s} - \lambda_1} - \frac{1}{\sqrt{s} - \lambda_2} \right), \quad (3)$$

where $\lambda_{1,2} = -(\tau_R^{1/2}/2\tau)(1 \mp \sqrt{1 - 4\tau/\tau_R})$ are the roots of the equation $s + (\tau_R s)^{1/2} \tau^{-1} + \tau^{-1} = 0$.

The inverse transform of (3) has the form [12]

$$v^*(t) = \frac{F}{M} \left\{ \tau + \frac{1}{\lambda_2 - \lambda_1} \sum_{i=1}^2 (-1)^i \frac{1}{\lambda_i} \exp(\lambda_i^2 t) \operatorname{erfc}(-\lambda_i \sqrt{t}) \right\}. \quad (4)$$

The behavior of this solution at $t \rightarrow 0$ is $v^*(t) \approx Ft/M$. At long times we have

$$v^*(t) \approx (F\tau/M) \left[1 - (\tau_R/\pi t)^{1/2} + \dots \right].$$

Using these formulas, the force F can be determined through the measured mean velocity $v^*(t)$ at any time t . At long times, due to the hydrodynamic aftereffect, this velocity depends on the relation τ_R/t and very slowly approaches the limiting value F/γ . The determined force is

$$F \approx \gamma v^* \left(1 + \sqrt{\tau_R/\pi t} \right), \quad \tau_R/t \ll 1. \quad (5)$$

The stochastic motion of the particle, for constant γ , does not influence the determination of the force, since its contribution to the drift velocity is zero. It is easy to see that our correction of the standard result for the force, $F = \gamma v^*$, can be significant. To demonstrate it, let us turn to the recent work [8]. The smallest observation times in this experiment were $\sim 10^{-8}$ s. At such times the measured force should be proportional to v^*/t . At longer times, when the ratio τ_R/t is small, (5) applies. For spherical particles $1 \mu\text{m}$ in radius, which are suspended in

water at room temperatures, we have $\tau_R \sim 10^{-6}$ s. Thus, at the times $\sim 10^{-5}$ s the correction represents almost 20% and slowly drops with the increase of time, approaching 1% of γv^* at $t \cong 3\text{ms}$.

III. IS THE BROWNIAN MOTION IN FLUIDS OF SUPER-DIFFUSIVE CHARACTER?

In the standard Langevin approach to the motion of a BP the friction force that the particle feels during its motion is the Stokes-Einstein force. When, however, the non-stationary motion of the particle in an incompressible fluid is studied, this approach fails and this resistance force should be replaced by the Boussinesq-Basset force that reflects the memory in the system or the so-called hydrodynamic aftereffect (see [10, 13] and references therein). Such a situation can be described by the generalized Langevin equation (GLE) (1) with $F = 0$. Now our aim is to calculate the velocity autocorrelation function (VAF) of the particle, $\phi(t) = \langle v(t)v(0) \rangle$, and its mean square displacement (MSD). For the motion along the direction x , $X(t) = \langle \Delta x^2(t) \rangle = \langle [x(t) - x(0)]^2 \rangle$. Here $\langle \dots \rangle$ stays for statistical averaging. Since the quantities $v(t)$ and $x(t)$ are stochastic variables, we do not use zero initial conditions, as it is often done for them [14]. Instead, assuming the initial equilibrium between the particle and the solvent, in agreement with the equipartition theorem for the particle of mass M , the condition $\phi(0) = k_B T/M$ is used for the VAF. Multiplying (1) by $v(0)$ and assuming that $\langle \zeta(t)v(0) \rangle = 0$, after the average one obtains

$$M \dot{\phi} + \gamma \phi + \int_0^t \Gamma(t-t') \dot{\phi}(t') dt' = 0. \quad (6)$$

In the Laplace transformation the solution for $\tilde{\phi}(s) = \mathcal{L}\{\phi(t)\}$ is

$$\tilde{\phi}(s) = \frac{k_B T}{M} \frac{M + \tilde{\Gamma}(s)}{\gamma + s[M + \tilde{\Gamma}(s)]}, \quad (7)$$

with $\tilde{\Gamma}(s) = \gamma \tau_R^{1/2} s^{-1/2}$. The inverse transform of (7) is found after expanding this expression in simple fractions. For $\lambda_{1,2}$, defined after (3), we have

$$\tilde{\phi}(s) = \frac{k_B T}{Ms^{1/2}} \left(\frac{C_1}{s^{1/2} - \lambda_1} + \frac{C_2}{s^{1/2} - \lambda_2} \right), \quad (8)$$

with $C_1 = (\lambda_1 + \tau_R^{1/2} \tau^{-1})/(\lambda_1 - \lambda_2)$, and C_2 obtained by exchanging the roots $\lambda_1 \rightleftharpoons \lambda_2$. The time τ_R is defined before (1). Using the inverse transform $\mathcal{L}^{-1}\{s^{-1/2}(s^{1/2} - \lambda)^{-1}\} = \exp(\lambda^2 t) \times \operatorname{erfc}(-\lambda \sqrt{t})$ [12], where erfc is the complementary error function, we find

$$\phi(t) = (k_B T/M) \sum_{i=1}^2 C_i \exp(\lambda_i^2 t) \operatorname{erfc}(-\lambda_i \sqrt{t}), \quad (9)$$

with the asymptotes

$$\phi(t) \approx (k_B T/M) (1 - t/\tau + \dots), \quad t \rightarrow 0, \quad (10)$$

$$\phi(t) \approx \frac{k_B T}{M} \left(\frac{\tau_R}{\pi t} \right)^{1/2} \left[1 - \frac{\tau_R}{2t} \left(1 - \frac{2\tau}{\tau_R} \right) + \dots \right], \quad t \rightarrow \infty. \quad (11)$$

The MSD is obtained as $X(t) = 2 \int_0^t (t-s) \phi(s) ds$. This equation is derived representing the distance a particle moves in time as an integral of its velocity, $x(t) - x(0) = \int_0^t v(s) ds$:

$$X(t) = \frac{2k_B T}{M} \frac{1}{\lambda_1 - \lambda_2} \left\{ \frac{\lambda_2}{\lambda_1^4} \left[\frac{4}{3\sqrt{\pi}} (\lambda_1 t^{1/2})^3 + \lambda_1^2 t + \frac{2\lambda_1}{\sqrt{\pi}} t^{1/2} + 1 - \exp(\lambda_1^2 t) \operatorname{erfc}(-\lambda_1 t^{1/2}) \right] - (\lambda_1 \rightleftharpoons \lambda_2) \right\}. \quad (12)$$

As $t \rightarrow \infty$, up to the first term decreasing with time we have

$$X(t) = \frac{2k_B T}{M} \left[\frac{4}{3} \left(\frac{\tau_R t^3}{\pi} \right)^{1/2} + (\tau - \tau_R) t + 2 \left(1 - \frac{2\tau}{\tau_R} \right) \left(\frac{\tau_R^3 t}{\pi} \right)^{1/2} - (\tau^2 - 3\tau\tau_R + \tau_R^2) + (\tau_R - 3\tau)(\tau_R - \tau) \left(\frac{\tau_R}{\pi t} \right)^{1/2} + \dots \right]. \quad (13)$$

Since $X(t \rightarrow \infty) \sim t^{3/2}$, the solution has a super-diffusive character. At short times the MSD shows the expected ballistic behavior, $X(t \rightarrow 0) \sim k_B T t^2 / M$.

However, is this unexpected result correct? Equation (9) and the following equations (11) – (13) significantly differ from those found in the literature for the VAF as a solution of (1) with $F = 0$ [15 - 18],

$$\phi(t) = \frac{k_B T}{M} \frac{1}{\lambda_1 - \lambda_2} \left[\lambda_1 \exp(\lambda_2^2 t) \operatorname{erfc}(-\lambda_2 \sqrt{t}) - \lambda_2 \exp(\lambda_1^2 t) \operatorname{erfc}(-\lambda_1 \sqrt{t}) \right], \quad (14)$$

which looks quite similarly to (9) but has an asymptote very different from (11):

$$\phi(t) \approx \frac{k_B T}{2M} \frac{\tau \tau_R^{1/2}}{\pi^{1/2} t^{3/2}} \left[1 - \frac{3}{2} \left(1 - \frac{2\tau}{\tau_R} \right) \frac{\tau_R}{t} + \dots \right], \quad t \rightarrow \infty. \quad (15)$$

The MSD that follows from (14) is

$$X(t) = 2D \left\{ t - 2 \left(\frac{\tau_R t}{\pi} \right)^{1/2} + \tau_R - \tau + \frac{1}{\tau} \frac{1}{\lambda_2 - \lambda_1} \times \left[\frac{\exp(\lambda_2^2 t)}{\lambda_2^3} \operatorname{erfc}(-\lambda_2 \sqrt{t}) - \frac{\exp(\lambda_1^2 t)}{\lambda_1^3} \operatorname{erfc}(-\lambda_1 \sqrt{t}) \right] \right\}, \quad (16)$$

with the long time behavior

$$X(t) \approx 2Dt \left\{ 1 - 2 \left(\frac{\tau_R}{\pi t} \right)^{1/2} + \frac{2}{9} \left(4 - \frac{M_p}{M_s} \right) \frac{\tau_R}{t} - \frac{1}{9\sqrt{\pi}} \left(7 - 4 \frac{M_p}{M_s} \right) \left(\frac{\tau_R}{t} \right)^{3/2} + \dots \right\}, \quad (17)$$

where $D = k_B T \tau / M$ is the Einstein diffusion coefficient of the particle. Only at $\tau_R \rightarrow 0$ we have the same terms $\sim t$ and the next constant term in (12) and (17). The characteristic times τ_R and τ are not independent; they are connected by the relation

$\tau_R / \tau = 9\rho / (2\rho_p + \rho)$, where ρ_p is the density of the particle. Thus, (9) - (13) describe the normal diffusion when the density of the solvent is much smaller than that of the particle.

However, the question remains how to get the full solutions (14) and (16), which are frequently repeated in the literature. For particles with ρ_p close to ρ the difference between our solutions and those from the literature can be significant. In all the mentioned papers (except [15]), zero initial condition for the particle velocity is assumed. For example, in [17] both the projection of the particle position vector on the axis x , $x(t)$, and its velocity, $v(t)$, are assumed zero at $t = 0$: $x(0) = v(0) = 0$. It is argued that within the approximation of small thermal displacements there is no loss of generality in choosing these trivial initial conditions. Similarly, one finds in [18] that the velocity of the particle is determined by its velocity at earlier times via backflow effects in the fluid, but it is assumed that the particle is at the equilibrium position $x(0) = 0$ and at rest for $t \leq 0$. The latter condition is used also in the work [16], where independently the same solution for the VAF and MSD as in [17] has been obtained in an elegant (but incorrect) way based on the linear response theory [19]. Implicitly, as will be shown in our further work, $v(0) = 0$ also in the recent works [14]. In our opinion, the simultaneous use of the conditions $x(0) = 0$ and $v(0) = 0$ is not consistent for random variables $x(t)$ and $v(t)$ (they “never” can be met together). Moreover, it is either assumed or it follows from the calculations in all the works that the equipartition theorem holds, i.e. that $\langle v^2(0) \rangle$ is nonzero. We will show that if the above incorrect assumptions are abandoned, the correct (normal) diffusion can be obtained only if the random force at $t > 0$ is correlated with the particle velocity at $t = 0$. That is, we must require that $\langle \zeta(t) v(0) \rangle$ is nonzero at $t > 0$.

IV. CONCLUSION

In conclusion, we have obtained an exact solution for the drift velocity of a Brownian particle in an incompressible fluid under the action of a constant force. In the particle motion the memory of the hydrodynamic type has been taken into account. The found velocity is proportional to the applied force but depends in a complicated manner on the time of observation t . At short times it is proportional to t and at long times it contains algebraic tails, the longest-lived of which being $\sim t^{1/2}$. Due to this the velocity very slowly approaches the limiting value F/γ . As a consequence, the force F can significantly differ from the value that would be extracted from the drift measurements neglecting the inertial effects, which is a standard assumption in the interpretation of such experiments. The presented method is equally applicable in the case of force linearly depending on the particle position. For nonlinear forces, first the open question about the choice of convention to be used in stochastic calculus should be resolved in the future work.

Although the presented study is directly connected to the recent experiments on colloidal particles in fluids [1 - 4, 8, 9], our approach can be used to describe apparently very different problems from various branches of science and engineering. For examples, such as superionic conductors or charge density waves, we refer to [20]. In our calculations the memory in the dynamics of the system vanishes in the time as $t^{-1/2}$. Other laws

of the memory loss can be accounted for by a change of the kernel $\Gamma(t)$ in (1). The observable quantities can have a different physical meaning (e.g., the drift velocity of a Brownian particle would correspond to the mean current in electric circuits) but the basic equations and the method of their solution remain applicable to a number of dissipative systems coupled to a thermal bath.

ACKNOWLEDGMENT

This work was supported by the Agency for the Structural Funds of the EU within the project NFP 26220120033, and by the grant VEGA 1/0370/12.

The authors wish to thank their supervisor Prof. V. Lisý for proposing the theme and for continuous help.

REFERENCES

- [1] G. Volpe, L. Helden, T. Brettschneider, J. Wehr, C. Bechinger: "Influence of Noise on Force Measurements", *Phys. Rev. Lett.* 104, 170602 (2010).
- [2] R. Mannella, P. V. E. McClintock, Comment on "Influence of Noise on Force Measurement", *Phys. Rev. Lett.* 107, 078901 (2011).
- [3] G. Volpe, L. Helden, T. Brettschneider, J. Wehr, C. Bechinger: "Reply to Comment on 'Influence of Noise on Force Measurements'", *Phys. Rev. Lett.* 107, 078902 (2011).
- [4] T. Brettschneider, G. Volpe, L. Helden, J. Wehr, C. Bechinger: "Force measurement in the presence of Brownian noise: Equilibrium distribution method vs. Drift method", *Phys. Rev. E* 83, 041113 (2011).
- [5] W. T. Coffey, Yu. P. Kalmykov, J. T. Waldron: *The Langevin Equation*, USA, New Jersey: World Scientific (2005).
- [6] J. M. Sancho: "Brownian colloidal particles: Itô, Stratonovich or a different stochastic interpretation", *Phys. Rev. E* 84, 062102 (2011).
- [7] Yu. L. Klimontovich: "Nonlinear Brownian motion", *Physics – Uspekhi* 37, 737 (1994).
- [8] R. Huang, I. Chavez, K. M. Taute, B. Lukic, S. Jeney, M. Raizen, E.-L. Florin: "Direct observation of the full transition from ballistic to diffusive Brownian motion in a liquid", *Nature Physics* 7, 576 (2011).
- [9] Th. Franosch, M. Grimm, M. Belushkin, F. Mor, G. Foffi, L. Forro, S. Jeney: "Resonances arising from hydrodynamic memory in Brownian motion - The colour of thermal noise", *Nature* 478, 85-88 (2011). See also Supplementary Material, doi:10.1038/nature10498.
- [10] V. Lisý, J. Tóthová, "On the (hydrodynamic) memory in the theory of Brownian motion", *arXiv:cond-mat/0410222* (12 pp.).
- [11] L. D. Landau, E. M. Lifshitz: *Fluid Mechanics*, England, Oxford: Pergamon (1987).
- [12] M. Abramowitz, I. A. Stegun, *Handbook of Mathematical Functions*, Washington, DC: National Bureau of Standards (1964).
- [13] J. Tóthová, G. Vasziová, L. Glod, V. Lisý: "Langevin theory of anomalous Brownian motion made simple", *Eur. J. Phys.* 32, 645 (2011); "Note on 'Langevin theory of anomalous Brownian motion made simple'", *Eur. J. Phys.* 32, L47 (2011).
- [14] F. Mainardi, A. Mura, F. Tampieri: "Brownian motion and anomalous diffusion revisited via a fractional Langevin equation", *Modern Problems of Statistical Physics* 8, 3 (2009); F. Mainardi, P. Pironi: "The Fractional Langevin Equation: Brownian Motion Revisited", *Extracta Mathematicae* 10, 140 (1996).
- [15] V. Vladimírsky, Ya. Terletsky: "Hydrodynamical theory of translational Brownian motion", *Zhur. Ekper. Teor. Fiz.* 15, 258 (1945) (in Russian).
- [16] Karmeshu: "Velocity fluctuations of charged particles in the presence of magnetic field", *J. Phys. Soc. Japan* 34, 1467 (1973).
- [17] E. J. Hinch: "Application of the Langevin equation to fluid suspensions", *J. Fluid Mech.* 72, 499 (1975).
- [18] H. J. H. Clercx, P.P.J.M. Schram: "Brownian particles in shear flow and harmonic potentials: a study of long-time tails", *Phys. Rev. A* 46, 1942 (1992).
- [19] R. Kubo: *Many Body Problems*, New York: W.A. Benjamin, Inc., p. 235 (1969).
- [20] A. Shit, S. Chattopadhyay, J. R. Chaudhuri: "Enhancement of transport properties of a Brownian particle due to quantum effects: Smoluchowski limit", *Chem. Phys.* 397, 48 (2012).

Laboratory complex for researching of the cooling systems for electric machines

¹Dmytro BORDUG (1st year), ³Martin BAČKO(3rd year), ⁴Tibor VINCE
SUPERVISOR: ²ANDRII GLADYR

^{1,2} Institute of Electromechanics, Energy Saving and Control Systems,
Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine

^{3,4}Dept. of Theoretical Electrotechnics and Electrical Measurement, FEI TU of Košice, Slovak Republic

¹gai@kdu.edu.ua ²martin.backo@tuke.sk ³jan.molnar@tuke.sk

Abstract—The laboratory complex for research of the cooling system for electric machines is presented. The physical model is proposed according to the specifics of design for ac motor. Experimental researches is produced with the aim of finding a dynamic characteristics of control object.

Keywords—Automatic tuning, forced cooling, temperature regulator.

I. INTRODUCTION

According to [1], in electric motors about 2/3 of failures are caused by failures of windings due to destruction of insulation. Long-life of insulation materials is determinate by its operation temperature. If its temperature is higher than nominal value the mean-time-between-failures of insulation will be extremely reduce and, as a result, the mean-time-between-failures of engine will be reduce too. That's why a problem of effective cooling for electric machines is actually. Review of researching about the control cooling system for power electric machines is

represented in [2]. For higher education it is necessary to create the equipment that allows researching the cooling system for electric machines in laboratory conditions.

II. DOCUMENT

For researching of the cooling systems it is necessary to have an electric machine that can operate in different modes which differ by level of thermal energy generated in active parts of engine. It is achieved by changing of load power with the help of second electric machine joined to first. Second machine has controlled shaft torque. Squirrel-cage motor in dynamic deceleration mode is selected as a load machine. Changing of resisting torque is realized by changing of direct voltage value applied to stator windings. It is needed to control speed of cooled machine in wide range for getting vent condition worse. It is realized by using of frequency changer.

The functional scheme has been given on Fig. 1.

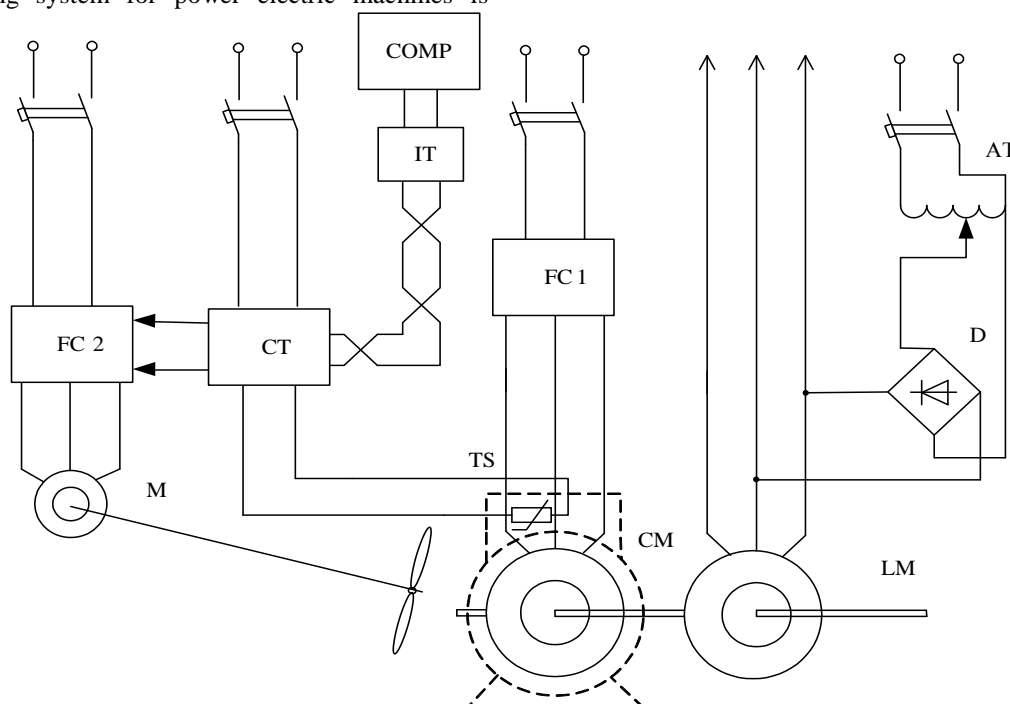


Fig. 1 The functional scheme of laboratory complex

Designed complex consist of cooled electric machine CM with frequency changer FC1; load machine LM with stator windings connected to the power source through autotransformer AT and rectifier D; ventilation machine constructed with ventilator V and drive motor M with its frequency changer FC2; controller of temperature CT with sensor TS; interface transformer IT for connecting to computer COMP.

Program for registration the experimental data had created in Labview environment. This program reads information from regulator with some frequency and creates graph, based on received data. Program also creates file at the beginning of registration. In this file information are saves as an electronic list. Besides that, developed program allows remote control. We can change set, run or stop temperature regulation.

The physical model of the cooling system for electric machines is proposed. This model has more simple construction and it is enough effective for organizing of researching. The offered physical model is represented on Fig. 2.

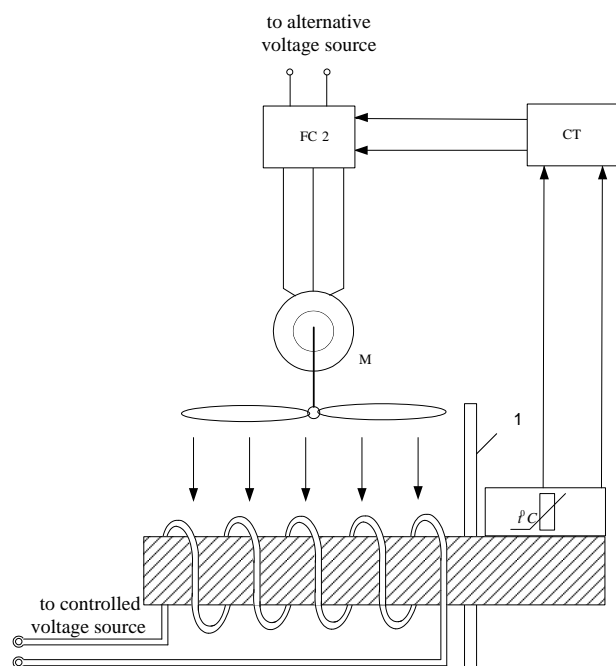


Fig. 2 The physical model of the cooling system for electric machines

Temperature is measured through metallic mass. Partition protects sensor of temperature from flow of air and it prevents a distortion of measured value.

Developed laboratory equipment allows authors to begin work on some perspective scientific directions. It has some interest to research an efficiency of automatic tuning of proportional-integral-differential regulator on real object.

The process of automatic tuning happens in such way. Digital regulator operates as a two-set regulator. It turns-on and turns-off the ventilator (Fig. 3). As a result, temperature is change in next way (Fig. 4). After the end of process of automatic tuning the device have calculated optimum values of coefficient of PID-regulator. It have calculated periodicity of master pulse and recommended value of “acceleration” of adjustment also.

The process of automatic tuning is represented on figures 3 and 4.

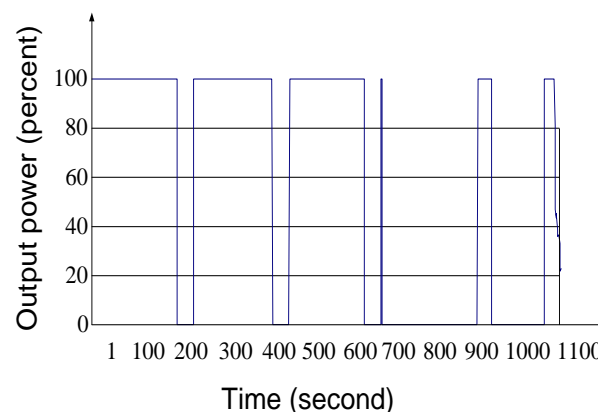


Fig. 3 The value of output signal of regulator

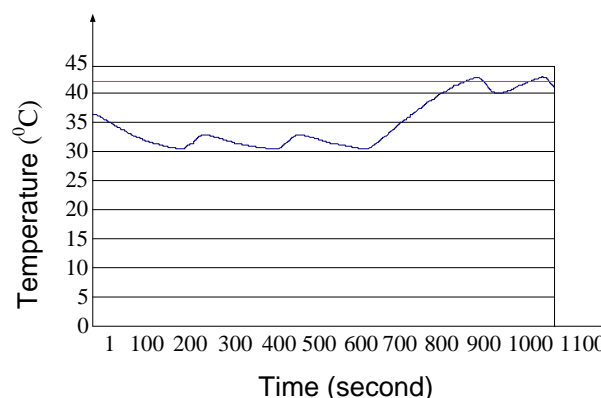


Fig. 4 The value of measured value of temperature

Next parameters of regulator were gotten thanks to automatic tuning:

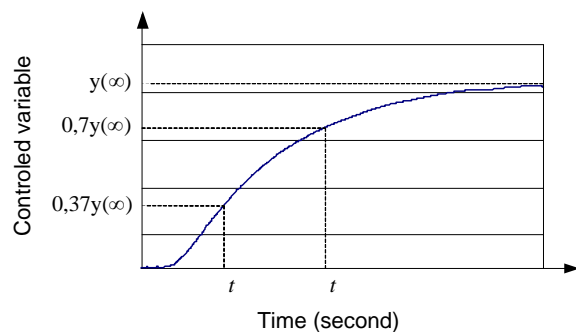
- proportional coefficient $k_p = 2.6$;
- integral coefficient $T_i = 87\text{sec.}$;
- differential coefficient $T_d = 13\text{sec.}$;
- period of master pulse $T_{\text{pulse}} = 1\text{sec.}$;
- filter constant $\tau_f = 0\text{sec.}$

Temperature transient process has been received using this parameters. Transient process has been imaged on figure 6.

Manual tuning was made for comparative analyses. At first it is necessary to make mathematical description of the device control object. According to [3], structure and parameters of control object can found with exchange real time response characteristic to equivalent characteristic. Equivalent characteristic consists of delay segment and exponent with time constant T_o . Approximated characteristic and real time response characteristic and real time response characteristic must be as similar as possible. Transfer function is described in equation (1)

$$W(p) = \frac{k_0 \cdot e^{-\tau \cdot p}}{T_o \cdot p + 1} \quad (1)$$

Approximation of time response characteristic can be more precise if it will make in next way. First of all moments of time will be determinate, when $y = 0,7 \cdot y(\infty)$ and $y = 0,33 \cdot y(\infty)$. This process is explain on figure 5.

Fig. 5 The determination of moments of time t' and t''

Then parameters of control object will be calculate according to (2).

$$\tau = 0.5 \cdot (3t'' - t'), \quad (2)$$

$$T_0 = 1.25 \cdot (t' - t'').$$

Coupling between output signal (y) and input signal (x) is determinate by amplification factor k_0 . Using (3), we can calculate the amplification factor.

$$k_0 = \frac{y(\infty)}{x(\infty)}. \quad (3)$$

If we have the mathematical description of the device control object, we can get to know a structure of correcting device (regulator). The coefficients can calculate by using (4).

$$k_p = \frac{0.95}{k_0 \cdot \tau / T_0}, \quad (4)$$

$$T_i = 2.4 \cdot \tau,$$

$$T_d = 0.4 \cdot \tau.$$

Next parameters of regulator were gotten thanks to manual tuning:

proportional coefficient $k_p = 9.17$;

integral coefficient $T_i = 191 \text{ sec.}$;

differential coefficient $T_d = 13 \text{ sec.}$

Temperature transient process has been received using this parameters. Transient process has been imaged on figure 7.

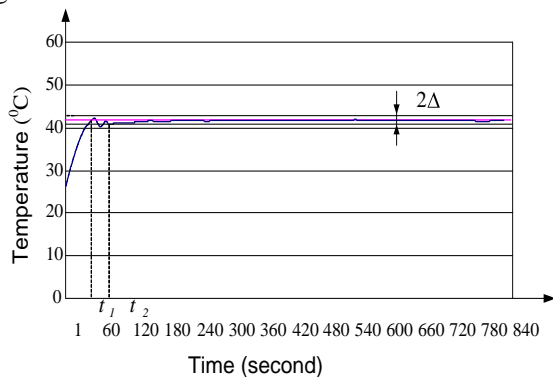


Fig. 6 Temperature transient process after automatic tuning

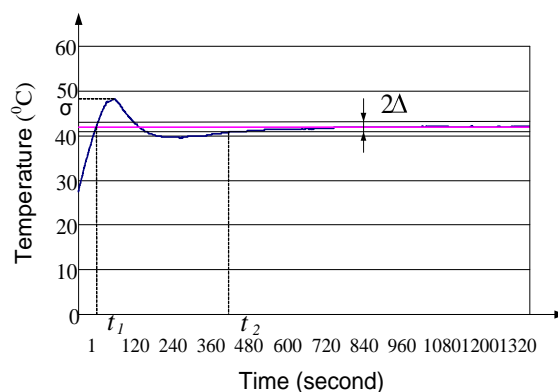


Fig. 6 Temperature transient process after manual tuning

Dynamic performances and static performances were determinate with the help of graphs of transient process.

System after automatic tuning has next performances: static error $\Delta T = 0\%$, overstate $\sigma = 0.5\%$, first achieve time $t_1 = 0.82$ minute, control time $t_2 = 1.3$ minute.

System after manual tuning has next performances: static error $\Delta T = 0\%$, overstate $\sigma = 14.8\%$, first achieve time $t_1 = 0.9$ minute, control time $t_2 = 7.5$ minutes.

It is clear that automatic tuning allows getting transient process with high speed of response and without oversteating.

III. CONCLUSION

Designed laboratory complex gives a great possibility for researching of the cooling systems for electric machines. For example, the efficiency of different type of tuning was analyzed. The result of experiments proves that automatic tuning allows high quality of control. Manual tuning showed worse results, because this method does not consider influence of filter and discreteness of master pulse. But often automatic tuning cannot be produce in manufacture conditions because of technological limits. In such case it is necessary to use manual tuning that allows getting acceptable quality of control.

REFERENCES

- [1] Е.М. Лопухин, Г.А. Семенчуков "Автоматизированное проектирование электрических машин малой мощности", Москва, Высшая школа, 2002. (in Russian).
- [2] Е. Nosach "Review of researching about the control cooling system for power electric machines", Electromechanical and energy systems, modeling and optimization methods. The 8th International conference of students and young researchers in Kremenchuk April 08–09, 2010. – Kremenchuk: KSU, 2010. – P. 245–247.
- [3] И.К. Петров "Курсовое и дипломное проектирование по автоматизации производственных процессов", Москва, Высшая школа, 1986. –с37-50 (in Russian).

Laboratory frame ^1H spin-lattice relaxation study of isotactic polypropylenes

¹Magdaléna UHRÍNOVÁ (2nd year), ²Peter DURANKA (2nd year)

^{1,2,3} Dept. of Physics, FEI TU of Košice, Slovak Republic
Supervisor: ³Dušan OLCÁK

magdalena.uhrinova@tuke.sk, peter.duranka@tuke.sk, dusan.olcak@tuke.sk

Abstract—Two samples of isotactic polypropylene, produced by different polymerizations were studied with the use of magic angle spinning (MAS) ^1H NMR spectroscopy in this work. First of them metallocene isotactic polypropylene (m-iPP) and the second Ziegler-Natta isotactic polypropylene (ZN-iPP) prepared in pelletized form were the subject of this study. Laboratory frame spin-lattice relaxation times T_1 (^1H) were determined in temperature range 30 – 190°C. This range includes glass transition relaxation and melting process of both polymers. The temperature dependence provided information on the motions of particular groups of iPP which are associated with glass transition relaxation and melting process. The spin-lattice relaxation times $T_{1,min}$ represent minima of glass transition relaxation and melting process. In both cases the minima were lower for m-iPP than for ZN-iPP and the experimental $T_{1,min}$ values are higher than theoretical values. For m-iPP the experimental values are in better relation with the theoretical ones. The differences between experimental $T_{1,min}$ values were explained by serious limitations of motions resulting from stronger entangled polymer chains in ZN-iPP than in m-iPP.

Keywords—Isotactic polypropylenes, spin-lattice relaxation time T_1 (^1H), glass transition relaxation, melting process

I. INTRODUCTION

Isotactic polypropylene is nowadays an important material often used in technical areas because of its properties formed thanks to its chains being crystallized in helical forms, which means it does have a crystal structure and physical characteristics that are strongly dependent on the preparation process [1]. Historically, it is the very first highly stereoregular synthetic polypropylene [2].

There are many situations when we need a certain powerful technique to be able to get detailed information about the dynamics, structure (three-dimensional) or topology of solutions and solids and its molecules. NMR (Nuclear Magnetic Resonance) spectroscopy is exactly a technique that offers such benefits. NMR techniques are very appropriate for purposes of analyzing morphology and molecular motion of studied materials, due to their high sensitivity. Therefore, they are widely used for studies of order characterization, orientation or alignment of molecular chains and their dynamics [2-4].

The spin-lattice relaxation times in the laboratory (T_1) and

in the rotating ($T_{1\rho}$) frames characterize the rate of establishing of equilibrium between a spin system and its surroundings in the respective frame [4, 5].

The spin-lattice relaxation time T_1 is sensitive to molecular motions of the frequency of the order of 10 - 100 MHz and the rotating frame relaxation time $T_{1\rho}$ to the slow motions with correlation frequencies in the range 10-500 kHz [3]. In motionally heterogeneous polymeric systems the observed relaxation is determined both by the intrinsic relaxation behaviour of the different regions and by the transport of magnetization. The last mentioned happens either due to spin-diffusion or due to significant diffusion of the polymer chains.

It is known that the temperature dependence of the spin-lattice relaxation time of iPP has two minima within the broad temperature range: a high temperature minimum related to a segmental motion of chains in the non-crystalline regions corresponding to the glass transition and a low temperature minimum connected with the rotation of the CH_3 groups [4].

To fulfill the goal of this work that was to analyze and evaluate influence of the preparation technology on the molecular dynamics of iPP, we performed measurement of ^1H spin-lattice relaxation times in laboratory frame.

II. EXPERIMENTAL

A. Materials

Two iPP samples in pelletized form prepared by different polymerizations are the subject of this study. The first of them known as TATREN HG 1007 (ZN-iPP) was prepared by the polymerization using classical Ziegler-Natta catalysis. Crystallinity and melting temperature of this sample, which were estimated by DSC technique, are 55%, 163.6°C, respectively, and glass transition temperature $T_g=10^\circ\text{C}$ was estimated from DMTA measurements. The other one known as METOCENE HM 562 N (m-iPP) was prepared by metallocene-catalysed polymerization. Degree of the crystallinity, melting temperature and glass transition temperature are 52%, 145.2°C and 12°C, respectively.

B. NMR measurements

All measurements were performed by Varian solid state NMR spectrometer. The ^1H resonance frequency was 400

MHz. The temperature range of the measurements was 30 – 190°C. The experiments were carried out with a probe-head using the 4mm rotor under the magic angle spinning conditions at the spinning rate of 10kHz. For the determination of T_1 inversion recovery method was used. Calculations were done using Varian VnmrJ 3.2 software.

III. RESULTS

Fig. 1 illustrates the spin-lattice relaxation process detected by inversion recovery technique. The time dependence in the Fig. 1 was obtained from the ^1H NMR spectra measured for m-iPP at 135°C.

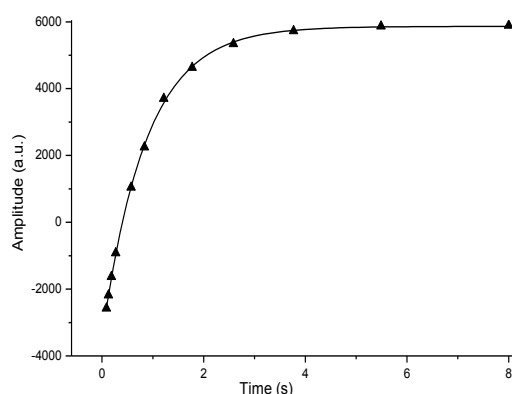


Fig. 1. The time dependence of the amplitude of the ^1H NMR lines detected during inversion recovery experiment performed for m-iPP at 135°C. The points represent the amplitudes of the ^1H NMR lines related to the CH_3 groups. The solid line shows the fitted exponential course of the spin-lattice relaxation process.

The Fig. 2 represents the temperature dependences of the spin-lattice relaxation times T_1 (^1H) for both samples m-iPP and ZN-iPP. The measurements were carried under magic angle spinning at frequency of 10 kHz in the temperature range from 30-190°C.

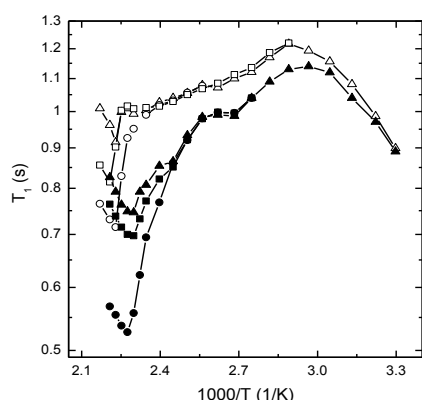


Fig. 2 The temperature dependences of the spin-lattice relaxation time T_1 of m-iPP (solid symbols) and ZN-iPP (open symbols) related to methine, methylene and methyl groups estimated by the peaks observed at 1.61 (■, □), 1.30 (●, ○) and 0.91 (▲, Δ) ppm, respectively

According to the previous papers, in the temperature range above 70°C three relaxation processes can be found. They are glass transition relaxation, α -relaxation process within crystalline domains and relaxation process associated with

melting. Using lower resonant frequencies for the spin-lattice relaxation time T_1 measurements performed on iPP [4, 6, 7] only glass transition relaxation process was observed within the temperature range between room and melting temperatures. For measurements reported in this paper the higher frequency (400 MHz) was used and from this reason the minimum related to the glass transition relaxation was shifted to the melting process. Then the overlapping of the glass transition relaxation with melting process could be observed in the temperature dependences of the spin-lattice relaxation times T_1 .

Fig. 2 shows that the spin-lattice relaxation related to the glass transition relaxation begins to be effective at 70°C. Due to the splitting of the spectra measured at temperatures above 70 °C, the values of spin-lattice relaxation times T_1 can be estimated within a broad temperature range for the protons giving peaks with chemical shifts of 1.61, 1.30 and 0.91 ppm, which we have assigned to methine, methylene and methyl protons.

The temperature dependences of the relaxation times T_1 related to the particular groups reach sharp minima at about the same temperature, 162-166 °C for m-iPP and 175-180 °C for ZN-iPP (Tab. 1). These minima are above the melting points of both samples.

For the ZN-iPP sample the temperature dependences of the spin-lattice relaxation times T_1 associated with the methyl and methine groups of the chains show broad minima at 162 °C. The dependence of the spin-lattice relaxation time T_1 for methylene groups shows the slope within the narrow temperature region about this temperature. The values of particular relaxation times $T_{1, \min}$ related to the minima occurring above the melting temperatures differ significantly from one another and differences were also found between relaxation times related to the same proton groups in different samples. (Fig. 2, Tab. 1).

TABLE 1
THE TEMPERATURES T_{\min} AND SPIN-LATTICE RELAXATION TIMES $T_{1, \min}$ RELATED TO THE POSITIONS OF THE MINIMA OF THE TEMPERATURE DEPENDENCES OF THE SPIN-LATTICE RELAXATION TIMES T_1 ASSOCIATED WITH THE MELTING PROCESS. THE THEORETICAL VALUES $T_{1, \min}$ FOR METHYLENE AND METHYL GROUPS WERE FOUND TO BE 0.31 AND 0.63 s, RESPECTIVELY. UNCERTAINTIES OF THE MEASUREMENTS WERE SET TO $\pm 0.1\text{K}$ FOR T_{\min} AND $\pm 0.01\text{s}$ FOR $T_{1, \min}$.

Proton group	m-iPP		ZN-iPP	
	T_{\min} (°C)	$T_{1, \min}$ (s)	T_{\min} (°C)	$T_{1, \min}$ (s)
CH_2	166	0.53	175	0.72
CH	162	0.70	180	0.82
CH_3	162	0.74	175	0.92

By comparing the glass transition temperatures (10°C and 12°C for ZN-iPP and m-iPP, respectively) of the studied samples it can be estimated, that the minima associated with the glass transition relaxations could be observed at about the same temperature. Two minima of the temperature dependences of the spin-lattice relaxation times are observed for the ZN-iPP. The temperature dependences for m-iPP show only one minimum at temperature of glass transition relaxation of ZN-iPP. It can be deduced from Fig. 2 that the

$T_{1, \min}$ for m-iPP is lower than $T_{1, \min}$ for ZN-iPP. The lower value of $T_{1, \min}$ for m-iPP can be also inferred from the fact that at the temperatures above 70 °C, the spin-lattice relaxation times T_1 for ZN-iPP are higher than those for m-iPP. Less restricted segmental motion in the amorphous phase of m-iPP than in that of ZN-iPP can be the reason of the relaxation time behavior as observed in the discussed temperature range [8].

The interpretation of the temperature dependences of the spin-lattice relaxation times related to the melting process is based on the assumption that the magnetic dipole-dipole interactions between equal nuclei is predominant relaxation mechanism and the isotropic reorientation of two-spin system with a single correlation time is the main motional process [9, 10, 11].

The satisfactory agreement was reached between theoretical and experimental values of the spin-lattice relaxation times $T_{1, \min}$ for m-iPP. The theoretical $T_{1, \min}$ values were calculated from the condition :

$$T_{1, \min} = \frac{\omega_0}{\langle \Delta\omega^2 \rangle} \quad (1)$$

,where ω_0 represents angular frequency and $\langle \Delta\omega^2 \rangle$ is second moment for isotopic rigid lattice.

The higher experimental $T_{1, \min}$ values as compared to the theoretical indicate that dipole-dipole interactions of methylene and methyl protons were not completely averaged in melting process. The interactions of these protons with the protons of neighbouring groups, which were not taken into account in our calculations, may be responsible for these differences.

The experimental $T_{1, \min}$ values related to the particular groups of ZN-iPP are obviously higher than those related to m-iPP (Fig. 2, Tab. 1). The differences between experimental and corresponding theoretical $T_{1, \min}$ values for ZN-iPP are larger than those for m-iPP. The comparison of the experimental $T_{1, \min}$ with theoretical for particular groups provides information not only on the motion of the individual groups but also on the polymer chain motions and more severe restrictions of motions of polymer chains in melted ZN-iPP as compared with those in m-iPP can be inferred from the $T_{1, \min}$ data. The reason of the different restriction can be possible entanglement of the high molecular weight chains in molten samples and stronger entanglement of ZN-iPP chains could give an explanation of their more severe restrictions of motion as compared with m-iPP chains. [2, 12, 13].

- [6] Kienzle, U., Noak, F., Von Schütz, J. Koll. Z. Z. Polym. 296 (1970) 129 - 137
- [7] McBrierty, V. J., Douglass, D. C., Falcone, D. R., J. Chem. Soc. Faraday Trans. II 68 (1972), 1051 – 1059
- [8] Tanaka, H., Saito, K., Colloid & Polymer Science, 266 (1998), 1 – 5
- [9] Bloembergen, N., Purcell, E. M., Pound, R. V., Phys. Rev., 73 (1948) 679
- [10] Hennel, J. W., Klinowski, J., Fundamentals of Nuclear Magnetic Resonance, Longmann Scientific and Technical, 1993, p. 221, 217
- [11] Kitamaru R., Nuclear Magnetic Resonance Principles and Theory, Elsevier, 1990, 122, 213
- [12] Arranz-Andrés, J., Peña, B., Benavente, R., Pérez, E. and Cerrada, M. L., Eur. Polym. J., 43 (2007) 2357 - 2370
- [13] Gómez-Elvira, J. M., Tiemblo, P., Elvira, M., Matisova-Rychla, L. and Rychly, J. Polymer Degradation and Stability, 85 (2004), 873 – 882

REFERENCES

- [1] T. Hatanaka: Polymer Degradation and Stability, 64, 313-319 (1999).
- [2] Busico, V. and R. Cipullo, Progress in Polymer Science, 26 (2001) 443 – 533
- [3] L. Ševčovič, L. Mucha: Solid State Nuclear Magnetic Resonance, 36, 151 (2009)
- [4] Olčák, D., Murín, J., Uhrin, J., Rákoš, M., Schenk, W., Polymer 26 (1985) 1455 – 1458.
- [5] M. Geppi et al.: Solid State Magnetic Resonance, 12, 15 (1998)

Mathematical Analysis of Harmonic Potential Field Method

¹Michal KALAVSKÝ (3rd year),
Supervisor: ²Želmíra FERKOVÁ

^{1,2}Dept. of Electrical Drivers and Mechatronics, FEI TU of Košice, Slovak Republic

¹michal.kalavsky@tuke.sk, ²zelmira.ferkova@tuke.sk

Abstract— Potential field method is one of the most using method for path planning of mobile robot. This method is easy using for static system where is known size of area, goal position and position of obstacles. Using for dynamic system is essentially more complicated because position of obstacles can change. Potential field method belongs to methods of planning on grid.

Keywords—field, potential field, harmonic function

I. INTRODUCTION

In recent years known are several methods for path planning of mobile robot. What does planning mean? Planning is possible to define as finding algorithm that is capable satisfy safe path from start position to the goal position. However, obstacles can be found in space, therefore algorithm must satisfy safe path without collision with obstacles. Finding of algorithm for path planning of mobile robot using potential field method is object of my study work therefore it will be described in this paper.

Potential field method is very well understandable method for path planning of mobile robot. Potential field method belongs to planning on grid. It means that real space is represented by grid. Imagine real space where robot can move. Real space is possible to divide on cells. Every cell represents certain section of real space and has his evaluation. By this process is achieved space where the grid introduces virtual space with finite number of cells. It is helpful for moving robot because his motion is controlled by evaluation of every cell.

Path planning using potential field method consists of two problems that are necessarily to solve. The first problem is generation of potential field that is represented real space. And the second problem is to find algorithm that is capable find safe path in potential field.

In this paper described will be the first problem, generation of potential field. Therefore, our study focuses on mathematical analysis of field, potential field and harmonic function.

II. FIELD, ENERGY, WORK AND POTENTIAL OF FIELD

A. Fields

In this section, we will be defined general term as field and, more specifically, potential field. In the following development we will be used the Cartesian coordinate system.

Field is possible to define in accordance to [2] as a set of functions of space and time. In general, we know two kinds of field. The first is material field and the second force field. Material fields describe some physical property (as temperature or magnetization) of material at each point of the material and at a given time. Force fields describe the forces that have effect at each point of space at a given time, for example magnetic fields induced by electrical currents.

Also, field is possible to class as scalar or vector. A scalar field is a single function of space and time. A vector field is characterized by three functions of space and time, by the components of the field in three orthogonal directions.

A vector field can be described by its field lines. Field lines are tangent to the vector field at every point. Small displacement along a field line must have x , y , z components proportional to the corresponding x , y , z components of the field at the point of its displacement. Therefore, if \mathbf{F} is a continuous vector field, so its field lines are described by integration of the differential equation in form:

$$\frac{dx}{F_x} = \frac{dy}{F_y} = \frac{dz}{F_z} \quad (1)$$

B. Energy and Work of the Field

Now, consider a particle that force field \mathbf{F} has effect on particle, see Fig. 1. This particle can be for example electrical charge that is moving under the influence of an electrical field.

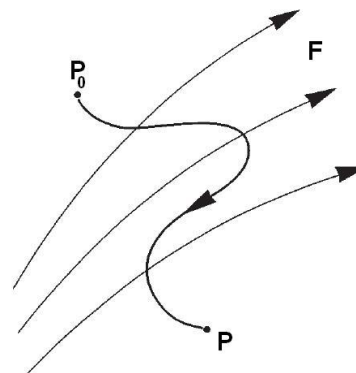


Fig. 1 Motion of particle from position P_0 to position P under the influence of force field \mathbf{F}

The kinetic energy expended by the force field on motion the particle from position P_0 to position P is described in accordance to [2] as the work made by the force field \mathbf{F} .

Second Newton's law of motion requires that momentum of the particle at any moment must change at a rate proportional to the magnitude of the force field and in a direction parallel to the direction taken by the force field at the location of the particle. It can be described as:

$$\lambda \cdot \vec{F} = m \cdot \frac{d}{dt} \cdot \vec{v} \quad (2)$$

where λ is a constant and \vec{v} is the velocity of the particle.

Let for this case is $\lambda=1$ and we multiply both sides of equation (2) by \vec{v} . Then it is obtained equation in form:

$$\vec{F} \cdot \vec{v} = \frac{1}{2} m \cdot \frac{d}{dt} \cdot \vec{v}^2 = \frac{d}{dt} \cdot E \quad (3)$$

where E is kinetic energy of the particle.

If motion of particle is from position P_0 to position P during time interval t_0 to t (see Fig. 1), then the change in kinetic energy is described as follows:

$$E - E_0 = \int_{t_0}^t \vec{F} \cdot \vec{v} \cdot dt = \int_{P_0}^P \vec{F} \cdot d\vec{s} = W(P, P_0) \quad (4)$$

where $d\vec{s}$ represents elemental displacement of the particle.

Magnitude $W(P, P_0)$ is the work that is needed for motion of particle from position P_0 to P . Equation (4) describes that the change in kinetic energy of the particle is equal the work made by force field \vec{F} .

In general, work needed on motion of particle from position P_0 to position P is different if trajectory of motion is different. In special case, if work is independent from trajectory of motion, then vector field is said to be conservative. Now, consider conservative field a move the particle an additional small distance Δx parallel to the x axis, see Fig. 2. Then equation (5) is obtained in form:

$$W(P, P_0) + W(P + \Delta x, P) = W(P + \Delta x, P_0) \quad (5)$$

what is possible to rewrite in form:

$$W(P + \Delta x, P_0) - W(P, P_0) = W(P + \Delta x, P) = \int_P^{P+\Delta x} F_x(x, y, z) dx \quad (6)$$

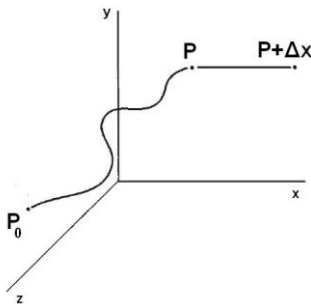


Fig. 2 Motion of particle from position P_0 to position P in conservative field

Integral in equation (6) can be solved by dividing both sides of the equation by parameter Δx . Then equation (7) is obtained in form:

$$\frac{W(P + \Delta x, P_0) - W(P, P_0)}{\Delta x} = F_x(X + \epsilon \Delta x, y, z) \quad (7)$$

where $0 < \epsilon < 1$. If Δx becomes arbitrarily small, then is obtained:

$$\frac{\partial W}{\partial x} = F_x \quad (8)$$

We can reiterate this derivation for other directions, for y and z directions. We can multiply each equation by corresponding unit vectors and add them to equation (8) and then new

equation is obtained:

$$\vec{F}(x, y, z) = \left(\frac{\partial W}{\partial x}, \frac{\partial W}{\partial y}, \frac{\partial W}{\partial z} \right) = \nabla W \quad (9)$$

Equation (9) says following implication: The derivative of the work in any direction is equal to the component of force in that direction. The vector force field \vec{F} is defined by the scalar field W , which we call the work function of vector force field \vec{F} . Therefore, we can say that conservative field is defined by the gradient of its work function. By equation (4), (9) is possible to show the opposite relationship. If work function W has continuous derivatives, then it is possible to integrate equation (9) as follows:

$$W(P, P_0) = \int_{P_0}^P \vec{F} \cdot d\vec{s} = \int_{P_0}^P \left(\frac{\partial W}{\partial x} dx + \frac{\partial W}{\partial y} dy + \frac{\partial W}{\partial z} dz \right) = \int_{P_0}^P dW = W(P) - W(P_0) \quad (10)$$

From equation (10) results that size of work depends only on the values at endpoints P and P_0 , not on the trajectory taken. This is exact definition of conservative field.

Now consider vector field that is described by equation (9) and has continuous derivatives. So we can say that is conservative. If trajectory of particle is a closed loop, then P equals P_0 what means that $W(P, P_0)=0$ and no work is needed for motion of particle around the closed loop. [2] [6]

C. Potential of Field

Now we can define a potential of field. The potential ϕ of vector field \vec{F} can be defined as the work function or as its negative. In accordance to [3] potential ϕ of vector field \vec{F} is defined as follows: If particles of like sign attract each other (gravity fields), then $\vec{F} = \nabla \phi$ and the potential is equal to work made by the field. If particles of like sign repel each other (electrostatic fields), then $\vec{F} = -\nabla \phi$ and the potential is equal the work made against the field by the particle.

Now consider a constant that will be added to potential ϕ . The constant have no effect on changing of potential ϕ what means that:

$$\vec{F} = \nabla \phi \quad (11)$$

This constant can be chosen generally so that approaches 0 at infinity. From this results that the potential ϕ at the point P is given as follows:

$$\phi(P) = \int_{\infty}^P \vec{F} \cdot d\vec{s} \quad (12)$$

From equation (12) results that the value of the potential ϕ at specific point P is not very important as the difference in potential between two separated points. [1] [2]

III. HARMONIC FUNCTION

Potential field method for path planning of mobile robot has one the basic weakness, creation of local minimum. Consider a virtual space that is represented a real space. Then consider goal position where robot would be arrived. Local minimum is position in the virtual space when robot arrives to the position that is not goal position but he is thinking that this position is the goal position. In this case robot will be stuck in local minimum so he is embedded in infinity cycle. He is not able to go away from this position.

One way how to eliminate problem with creation of local minimum is using harmonic potential field method. The basic idea of this method consists in creation of only one minimum in the virtual space, global minimum represented goal position. If a goal position is global minimum, other minimum not exists in the virtual space and robot will be always arrived to the goal position.

Therefore, harmonic function is explained and described in other section of this paper.

In previous section of this paper conservative vector field \vec{F} was defined as scalar potential φ that has form $\vec{F} = \nabla\varphi$ or $\vec{F} = -\nabla\varphi$. If it is valid that $\vec{F} = \nabla\varphi$ and the vector field \vec{F} is conservative, then the vector field \vec{F} is termed as potential field. Potential field has his properties. One of them is described in the following: The potential φ of the vector field \vec{F} , under certain conditions, satisfies an important second-order differential equation called Laplace's equation that is in form:

$$\nabla^2\varphi = 0, \quad (13)$$

at points not occupied by sources of \vec{F} .

Now consider a situation that is depicted on Fig. 3. You can see a stretched rubber band subject to a static force directed in the y direction. The displacement φ of the rubber band in the y direction is described by the differential equation in form:

$$\alpha \frac{d^2\varphi}{dx^2} = -F(x), \quad (14)$$

where α is a constant and $F(x)$ is a force directed in the y direction per unit length in the x direction. If $F(x)=0$, then the rubber band lies along a line (see Fig. 3 (B)) and it is obtained equation in form:

$$\frac{d^2\varphi}{dx^2} = 0 \quad (15)$$

Note that it is valid for trivial one-dimensional case. The second-order derivative of a function is a measure of the function's curvature and the equation (15) indicates the obvious result: The stretched rubber band has no curvature in the absence of external forces. Laplace's equation (15) is not satisfied along any part of the band obtaining a local minimum or maximum. From previous statement results that if $\varphi(x)$ is to satisfy Laplace's equation (15), then for situation depicted on Fig. 3 (B) the maximum and minimum must occur only at the end points of the rubber band.

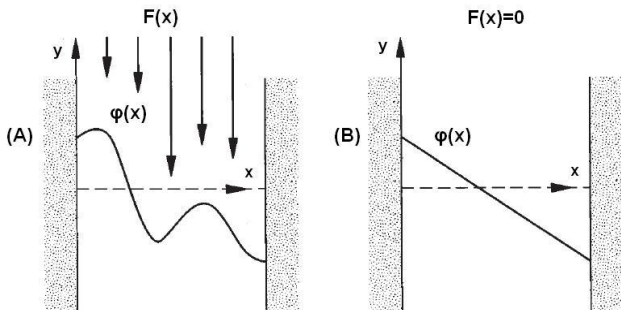


Fig. 3 Displacement in the y direction of stretched rubber band due to an applied force $F(x)$. (A) Static force $F(x) \neq 0$ that is evocating curvature along the x axis. (B) Static force $F(x) = 0$ so that $\varphi(x)$ has minimum and maximum only at the endpoints of the band.

Laplace's equation (15) can be extended to two- and three-dimensional cases. In these cases, the equations have form:

$$\frac{d^2\varphi}{dx^2} + \frac{d^2\varphi}{dy^2} = 0 \quad \text{and} \quad \frac{d^2\varphi}{dx^2} + \frac{d^2\varphi}{dy^2} + \frac{d^2\varphi}{dz^2} = 0 \quad (16)$$

From previous statements we can define in accordance to [2] harmonic function as follows: Harmonic function is function that satisfies the following criteria:

1. Satisfy Laplace's equation $\nabla^2\varphi = 0$.
2. Have continuous, single-valued first derivatives.
3. Have second derivatives.

From previous declaration we can say that harmonic function throughout a region R must have maximum and minimum on the boundary of region R and none within R itself. This is very useful for harmonic potential field method because this method must have only global minimum and maximum and none local minimum and maximum. Harmonic function is exactly basis of the harmonic potential field method.

Now consider one-dimensional case. The definition of the second derivative of harmonic function demonstrates another important property of a harmonic function. The second derivative of harmonic function $\varphi(x)$ is given as follows:

$$\lim_{\Delta x \rightarrow 0} \frac{1}{\Delta x^2} \left\{ \varphi(x) - \frac{1}{2} [\varphi(x - \Delta x) + \varphi(x + \Delta x)] \right\} = -\frac{1}{2} \cdot \frac{d^2\varphi}{dx^2} \quad (17)$$

If φ satisfies Laplace's equation, then the right-hand side of equation (17) vanishes and it is obtained equation:

$$\varphi(x) = \frac{1}{2} \lim_{\Delta x \rightarrow 0} [\varphi(x - \Delta x) + \varphi(x + \Delta x)] \quad (18)$$

From equation (18) results that the value of a harmonic function φ at any point is the average of φ at its neighboring points. This statement is suggested familiar property of a potential: A function can have no minimum or maximum within a region in which it is harmonic. [2] [4] [5] [7]

Find a solution of Laplace's equation is solving boundary values problem (Dirichlet type). In this case, it is needed to find formulation φ for entire region R where within region R must satisfy Laplace's equation $\nabla^2\varphi = 0$ and specific values are given for boundary of region R . This case is equal as harmonic potential field method case where it is needed to create a harmonic potential field with minimum and maximum on its boundary and within region R it satisfies Laplace's equation $\nabla^2\varphi = 0$.

IV. CONCLUSION

This paper describes potential field method for path planning of mobile robot. This paper is presenting mainly mathematical basis of potential field method, especially harmonic potential field method. For this method properties and conditions of harmonic function described above are very important. Exactly principle of harmonic function is also principle of harmonic potential field method.

Harmonic potential field method is often using for path planning of mobile robot. This method is applicable for static or dynamic system. However, application for dynamic system is more complicated because position of obstacles can be changed in every time.

ACKNOWLEDGMENT



This paper was developed with support of the Project “Centrum excelentnosti integrovaného výskumu a využitia progresívnych

materiálov a technológií v oblasti automobilovej elektroniky” (Centre of Excellence of Integrated Research and Exploitation the Advanced Materials and Technologies in the Automotive Electronics), ITMS 26220120055, that is co-financed from Structural Funds EU ERDF within Operational Program Research and Development OPVaV-2009/2.1/03-SORO and preferred axis 2 Support of Research and Development (70%).

REFERENCES

- [1] J. C. Latombe, Robot Motion Planning. Massachusetts, USA, Kluwer Academic Publisher, 1991. 651 pp. ISBN 0-7923-9129-2.
- [2] J. R. Blakely, Potential Theory in Gravity and Magnetic Applications. Cambridge, United Kingdom: Cambridge University Press, 1996. 441 pp. ISBN 0-521-57547-8.
- [3] O. D. Kellogg, Foundations of Potential Theory. New York, USA: Dover Publications, 1953. ISBN 0-486-60144-7.
- [4] I. C. Conolly – J. B. Burns – R. Weiss, Path Planning Using Laplace's Equation. Proceedings of the 1990 IEEE International Conference of Robotics and Automation. 13-18 May 1990, Cincinnati, USA. 2102-2106 pp. ISBN 0-8186-9061-5.
- [5] Ch. Shi – M. Zhang – J. Peng, Harmonic Potential Field Method for Autonomous Ship Navigation. Proceedings of the 7th International Conference on Intelligent Transport Systems Telecommunications. 6-8 June 2007, Sophia Antipolis, France. ISBN 1-4244-1178-5.
- [6] J. Kvasnica, Teorie elektromagnetického pole, Praha, Czech Republik: Academia Praha, 1985. 449 pp.
- [7] J. Vaščák, Využitie potenciálových polí v navigácii mobilných robotov, Technical University in Košice, 2008

Measuring of Availability and Reliability of FSO links from Measured Weather parameters

¹Matúš TATARKO (1st year)
Supervisor: ²Luboš OVSENÍK

^{1,2}Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

¹matus.tatarko@tuke.sk, ²lubos.ovsenik@tuke.sk

Abstract—The basic topic of this paper is focused on optical system, which is called Free Space Optics (FSO). It is the last mile optical system, which gives us free hand for installation. We don't need any special technology for installing. Establishing of this connection is very easy. The transporting medium is air, so there are many weather phenomena which have influence on the quality of connection. Visibility has the main influence on the quality of services. Thus this paper gives us a description of the device, which measures weather conditions such as density of fog, relative humidity and temperature. From measured data of density of fog is possible to calculate a visibility, which is the main factor for measuring availability and reliability of FSO links.

Keywords— fog sensor, FSO system, reliability and availability of services, visibility.

I. INTRODUCTION

The FSO system is growing up technology, which offers full duplex and usually protocol independent data transmission between two static points. Line of sight is needed [3,6]. Distance between this two points, (heads of FSO), is from several hundred meters up to few kilometres. Data rates are from hundreds of megabits up to tens of gigabits. Installation of this optical system is fast, easy and not very expensive. We don't need expensive fibre optics cable; rooftop installation and spectrum license are not required too.

Disadvantages of FSO systems are variations of receiving signal due to atmospheric phenomena, which depend on quality of air and line of sight. This fact limits the availability of FSO for a given transmission range, so it is very important to measure and know the weather properties such as fog parameters, humidity and temperature. FSO beams are absorbed and scattered by the air molecules as well as by the solid and liquid particles diffused in atmosphere. Absorption of the signal causes a decrease in signal strength. Scattering does not cause a decrease in signal strength, but it sends off the signal in different directions [1,6].

In following chapters we can find a detail description of device for measuring density of fog, temperature and humidity. It is called Fog sensor. In next part, there is a way, how to calculate values of visibility from measured data of density of fog. In final chapter, there are two experiments and their results. We choose one foggy day and one day without fog.

II. FOG SENSOR DEVICE

In Optoelectronic Laboratory at the Technical University in Košice, there is situated a device, which is called Fog Sensor and which is used for measuring weather conditions which are mentioned above such as density of fog, temperature and relative humidity (Fig.1). It measures data every second, 24 hours per a day of each of three variables. These three parameters are important for subsequent static and statistical evaluation of the quality of FSO communication environment. A fog has the main influence of FSO transmitting quality, which composes of water vapour or water droplets with 100 nm in diameter. In order to estimate the attenuation due to fog, content of liquid water is measured in unit (g/m³). This task usually requires relatively complex and expensive measuring instruments.

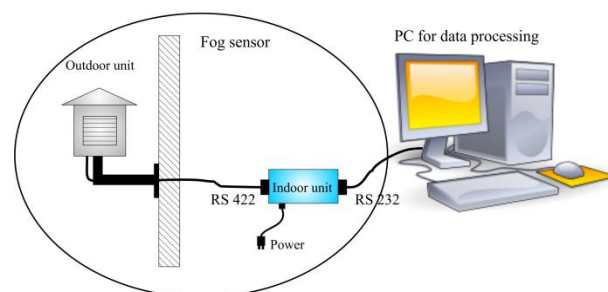


Fig. 1. Block scheme of Fog sensor

The Fog sensor consists of two parts. First part is called Outdoor unit, which performs the measurement. This unit is connected to another part, Indoor unit, with RS- 422 line. The Indoor unit is connected to PC through RS- 232 line. Obtained data are processed in PC in several different ways [3]. They can be saved to a file, rendered to chart or stored on server. All of these values are pre-processed in simulation program Matlab that allows us to obtain other parameters, which describe availability and reliability of FSO link. This parameter is marked like LWC and it means liquid water content of the atmosphere.

LWC is one of the most important parameters of the fog, which describes the mass of water drops in the volume units of the fog. It is measured in [g/m³]. Direct method for measuring LWC consists of extracting a known volume through a cotton pad or of rotating cups in an impeller apparatus, both to be weighed. Also changes of the electrical

resistance can be measured with a hot wire probe attached to an aircraft flying through the clouds. The value of LWC in the fog varies in a wide range. Fog attenuation can be calculated from LWC values [7].

III. DATA PROCESSING IN MATLAB

From fog sensor measured data we can calculate LWC parameter too, but we have to pre-process the measured data from the first column of source data in Matlab (TABLE I).

TABLE I
SOURCE DATA FROM FOG SENSOR BEFORE PROCESSING

Density	Temperature	Humidity	Averaging	Relative Time
227	4006	3121	100	0
213	4007	3121	100	1
213	4007	3121	100	2
218	4006	3122	100	3
211	4008	3121	100	4
216	4008	3121	100	5
216	4010	3121	100	6
216	4009	3121	100	7
258	4010	3121	100	8
215	4008	3121	100	9
218	4009	3121	100	10
214	4009	3121	100	11
214	4010	3121	100	12
215	4009	3121	100	13

Source data from fog sensor are just numbers without units. E.g. in a column of temperature, number 4006 represents some value of temperature in °C. For calculation values of fog density we use a source data from first column of TABLE I to equation (1).

$$F = (D_i \cdot \frac{5}{1024} - 0,96) \cdot \frac{0,5}{2,9} \quad [g/m^3] \quad (1)$$

where F is a density of fog and D_i is fog sensor output value (it is value from first column of TABLE I). If we want to get exact value of liquid water content, we need to define a constant which helps us to convert value of fog to value of LWC. This constant is marked as C and it is gains from ratio between the average liquid water content values W_i and the fog sensor output values D_i .

$$C = \frac{\sum_{i=1}^n W_i}{\sum_{i=1}^n D_i} = 0,7384 \quad [g/m^3] \quad (2)$$

where n is number of samples during whole fog event. This constant is used to convert the momentary sensor values D_i to momentary LWC values [2]:

$$LWC = F \cdot C \quad (3)$$

The visibility can be calculated from LWC. For the calculation we use the empirical formula for fog visibility as a function of fog density:

$$V = d \cdot (LWC)^{-0,65} \quad (4)$$

where V is a visibility in [km] and parameter d (without unit) takes on specific values for different fog conditions as shown in TABLE II below [5]:

TABLE II
TYPICAL VALUE OF PARAMETER “ d ” FOR DIFFERENT TYPES OF FOG

Type of fog	d
Dense haze	0,013
Continental fog (dry and cold)	0,034
Maritime fog (wet and warm)	0,060
Dense haze and selective fog	0,017
Stable and evolving fog	0,024
Advection fog	0,02381

We have mentioned a stable and evolving fog in this case. The stable air mass with cloud cover during the day, clear skies at night, light winds and moist air near the surface. These conditions often occur with a stationary, high pressure area [4].

Thus parameter d has value which corresponds with stable and evolving fog. So visibility is given by the equation below:

$$V = 0,024(LWC)^{-0,65} \quad (5)$$

IV. EXPERIMENTS AND RESULTS

In campus of Technical University in Košice two FSO systems are planned. In Fig. 2, there is a map of campus and there we can see possibility how to install these systems. First connection will be between Laboratory of Optoelectronics System (LOS) and the main building of Technical University in Košice (TUKE). Second connection will be between LOS and Institute of Computer Technology (ICT).

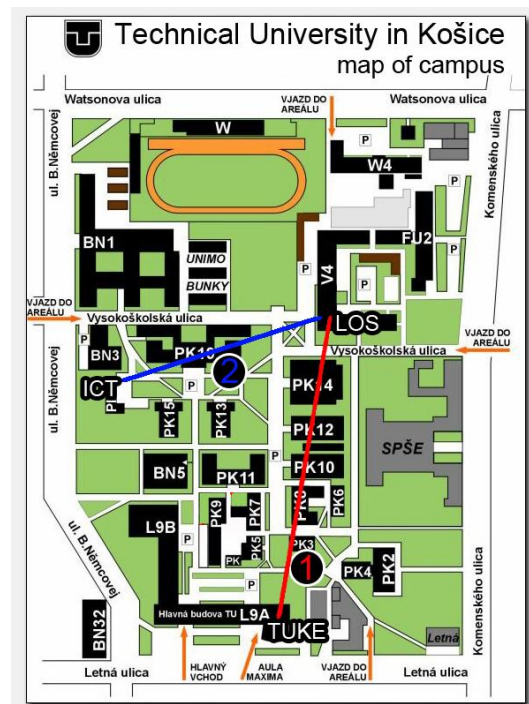


Fig. 2. Map of campus TUKE with FSO links

These two types of FSO links will be used for connections. Both FSO links will use systems from FSO companies. First system is from FSONA model SONAbeam 155E (red line).

Second system will use model Flightstrata 155-E, from Lightpointe Company (blue line). In next table, there are parameters of these two systems (TABLE III).

TABLE III
PARAMETERS OF FSO SYSTEMS

Parameters	FSONA SONAbeam 155E	Lighpointe Flightstrata 155- E
Wavelength	1550 nm	850 nm
Tx Power	100 mW	160 mW
Rx Power	-30 dBm	-30 dB
Rx Diameter	10 cm	8 cm
Directivity	2 mrad	2 mrad

When these parameters we give to the FSO System Simulator, we obtain a maximum distances in which the links are available. For Lightpointe system this distance is 1190 m and for FSONA system it is 1592 m. But these distances are for ideal conditions without any unfavourable weather effects.

In our case we mention these connections: first link will have 451m (FSONA system) and second link will have 312 m (Lightpointe system), (Fig.2). From these distances we can say, that they are available also at minimum visibilities too. But all values of visibility which are lower than the threshold distance, FSO systems will have results in loss of line.

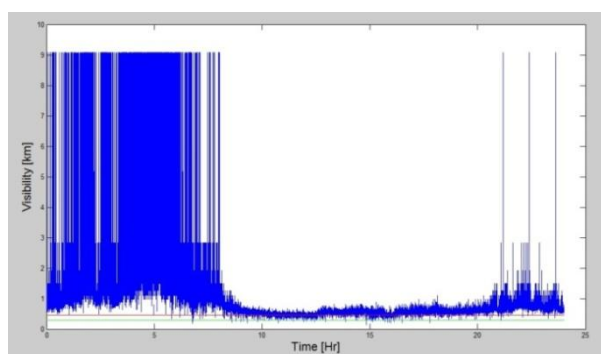


Fig. 3. Graph of visibility on 31th October 2011

From source data of fog sensor we can calculate an actual visibility in every second per day. In Matlab we can specify the threshold values of distances for both FSO systems. Graph of visibility is created with both thresholds (Fig. 3). For each system, program has created text file with distances which are lower than threshold.

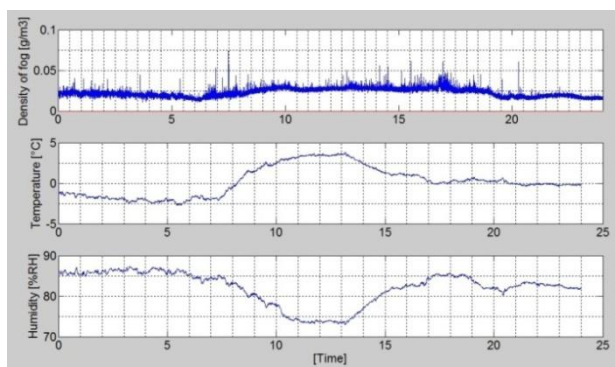


Fig. 4. Graphical output of 15th November 2011

In Fig. 3 red threshold belongs to FSONA system and green threshold belongs to Lightpointe system. From first look we can see that better availability has Lightpointe system. It is true of course, but we know exactly how many

percent per a day was link down. In this case from text files we know, that first link was 3004 times per a day down. It is 3,471% of day. Second link was down only 90 times per a day. It is 0,104 % of day. Measurement was performed in a foggy day 31th October 2011. We can calculate that availability of first FSO link was 96,539 % and 99,896 % for second FSO link.

Another day is illustrated in Fig. 4. This data was measured on 15th November 2011.

Graph of visibility is illustrated in Fig. 5.

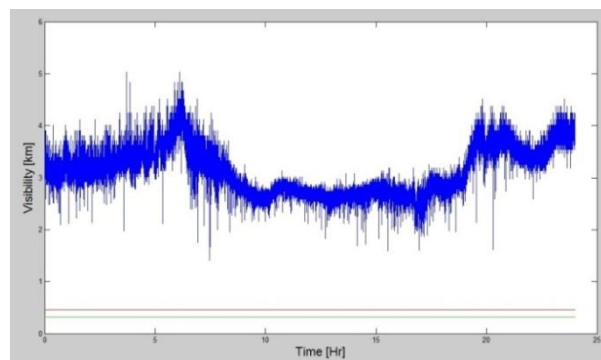


Fig. 5. Graph of visibility on 15th November 2011

From these graphs we can see that all values of visibility are higher than thresholds for our systems. So availability of both systems was 100 %.

V. CONCLUSION

This paper gives us simulation and analysis of FSO links, which are depended on the weather conditions, especially on fog. There are described a relation between liquid water content and visibility. Our Fog sensor gives us this information. From measure data we can statistically evaluate the availability and reliability of FSO links for certain period. It is very good knowledge for next installation of these systems in our campus.

ACKNOWLEDGMENT



Európska únia



Ministerstvo školstva, vedy, výskumu a športu SR

We support research activities in Slovakia / Project is co-financed from EU funds. This paper was developed within the Project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky", ITMS 26220120055.

REFERENCES

- [1] I. I. Kim, E. Korevaar "Availability of free space optics (FSO) and hybrid FSO/RF systems," *Proc. of SPIE*, vol. 4530, pp. 84-95, 2001.
- [2] Z. Kolka, O. Wilfert, V. Biolkova "Reliability of Digital FSO links in Europe," *Int. J. Electronics, Communications, and Computer Engineering*, vol. 1, no. 4, pp. 236-239, 2007.
- [3] L. Ovseník, J. Turán, P. Mišenčík, J. Bitó, L. Csurgai-Horváth "Fog density measuring system", *Acta Electrotechnica et Informatica*, 2012, pp. 1-5.
- [4] M. Reyman, J. Piasecki, F. Hosein and col. *Meteorological Techniques*, July 1998.
- [5] S. Sheikh Muhammad, M. Saleem Awan, A. Rehman. PDF Estimation and Liquid Water Content Based Attenuation Modeling for Fog in Terrestrial FSO Links. *RADIOENGINEERING*, Vol. 19, No. 2, June 2010.
- [6] S. Sheikh Muhammad, B. Flecker, E. Leitgeb Characterization of fog attenuation in terrestrial free space optical links. *Optical Engineering*, June 2007.
- [7] V. Zhurbenko *Electromagnetic Waves*, June 2011, ISBN 978-953-307-304-0, pp. 158-169.

MHP application in interactive TV

¹*Rastislav KOKOŠKA (5st year)*

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

¹rastko@pobox.sk

Abstract— In the article we have targeted services in the IPTV field, and more concretely MHP applications. The contribution provides an overview of possible types of services, their advantages and attributes. MHP services in interactive TV describe the architecture, layers, profiles, usage and application of MHP. Possible types of MHP enable different interactivity of IPTV utilization, e.g. in LCD TV receivers, which are mentioned at the end of the article. The above mentioned MHP services have several advantages, the advantage of interactivity of watching TV. The interactivity is also used in specialized receivers called Set Box. Among the other services required by final users belong the services like EPG, fast access to the Internet, and different multimedia services.

In the future, as well as in the present, the services are increasingly employed and increasingly required by providers, but final users as well.

Keywords— IPTV Internet Protocol TV, MHP –Multimedia Home Platform, Digital television, Interactive television, VAS – Value Added Services, EPG, Application Information Table (AIT), JAVA VM (JAVA Virtual Machine), API (application programming interface), DVB-J, Digital Storage Media – Command and Control (DSM-CC), DVB-S – satellite broadcasting, DVB-T – terrestrial broadcasting, DVB-C – Cable broadcasting, Pay Per View – paid services, Video on Demand, GEM – Globally Executable MHP.

I. INTRODUCTION

In the article, I will introduce the theoretical basis of the MHP services utilization in IPTV, and the other provided services. The reason of my interest is mobility of services, not only classical watching of TV programs, but interactivity of TV programs, watching and creativity of the possible services selection. At the end of the article, the possibilities of practical implementation into already existing and new appliances like TV sets and set top boxes (satellite and cable digital receivers) are introduced.

A. iTV and IPTV services

ITV is a television with interactive content. It combines watching of traditional television with communication through a reverse channel, mostly by means of a computer net (e.g. internet). IPTV (Internet Protocol TV), which uses already existing data net, mostly private – not public – for distribution of television and video (video on demand) services. For subscribers thus in ‘one package’ and with only one connection telephone, television and on-line services from one provider are available. Accordingly, IPTV can include also

what we refer to as iTV. Video on demand can be a subset of iTV (1). Interactive services are one of the strongest invitations delivered by digital television. The term ‘interactivity’ itself, for a viewer, means a possibility to change the contents of the screen. A typical representative is teletext [12].

B. Full interactivity services

Reaching full interactivity of services means the need of bi-directional communication between user and provider of the service. For creation of the reverse channel in present-day set top boxes, technology of Ethernet and integrated telephone modem are used mostly. Ethernet enables integration of set top boxes into a local network, and consequent usage of any other connections to the Internet (ADSL, ISDN, 802.11). User, an owner of a set-top box with a reverse channel, will be able to receive and send e-mail messages, browse websites, run the Internet shopping, download films, play computer games, etc. [12]. Possibilities of picture information in interactive television are strikingly broadened, and they increase its functionality. Its contents could be divided into two main groups. One is formed by TV programs and formats, the other one by Value Added Services – VAS (Tab.1) [2].

Value Added Services can relate to a program, but can be autonomous too. The most common type of VAS are informative (including electronic program guide – EPG), communicative, entertaining, transactional services, and generally interactive TV programs. From the being prepared data services: online voting during series, online shopping, home banking, choice of camera during sports broadcasting, interactive games, web browsing, full graphic teletext – superteletext, Pay per View, Video on Demand, etc.

Pay per View – with the paid series, there is a concept that within cumulative stream of DVB-T, certain series will be broadcasted repeatedly (in a loop circuit), and viewer, by means of reverse channel (used bi-directionally), will require the data necessary for its decoding, and will pay for the service. Then he can watch the series directly (as a stream), or save it on a local hard disk and watch it later, eventually several times – according to the rights he has bought.

Video on Demand – Service Video on Demand enables viewer to arrange his own program. If he selects from the transmitted data stream, they can be short video sequences, e.g. replayed shots during watching sports events. More extensive and more specified contents (full television series or feature-length films) will be then broadcasted by means of

reverse channel to the viewer who has chosen it [7].

Today, there are a lot of different contests and shows which use SMSs as a reverse channel for interaction with viewers. The content of the SMSs is used for viewers' voting (it influences the course of a contest or entertainment program), or is displayed on the screen as a chat contribution. Viewer, through his answers, can take part in a contest while being at home, in front of his TV set, he can bet on-line during a sports contest or race.

Interactive TV advertising (i-advertising) offers different ways of interaction between viewer and the Internet. It can be a request for additional information, or a testing sample by means of reverse channel, participation in different competitions arranged by performers, eventually mediation of a feedback on them, but also on-line shopping (T-commerce).

An integral part of iTV possibilities is electronic communication between viewers, usage of electronic mail and surfing on the Internet.

TABLE I

TYPES OF SERVICES WITH ADDED VALUE AND THEIR USAGE IN IDTV (2)

Type of Service VAS	Program equivalent VAS	Autonomous VAS
Electronic program guide (EPG)		On the basis of set-top box, countrywide or for a certain channel
Information	Extended news, traffic information, weather forecast, sport information live	Presentation of travel destinations in digital teletext
Interactive TV programs	See lines communication, entertainment, transaction	
Communication	Virtual discussion groups (chat by means of set-top box, mobile phone, Internet)	e-mail, instant messaging
Entertainment	Contests (participation in a contest or quiz from home), voting, ringing tone and picture download, etc.	Games, quizzes, etc.
Transaction	TV shopping, betting	Concert tickets, travel tickets, bank services.

All mentioned above creates not only technological chain, but also chain of values, whose yield is naturally high, and in a great deal depends and will depend on the will and interest of customers. The route of iTV contents from different sources to the customer is illustrated in Fig.1 [3].

Interactive television broadens reception and processing of information into multisource and multiappliance ones, which means usage of not only a TV set. As usage of iTV is something between watching TV and work with computer, whilst these two activities are converging and joining to each other, it needs continuous development of new hardware and software platforms.

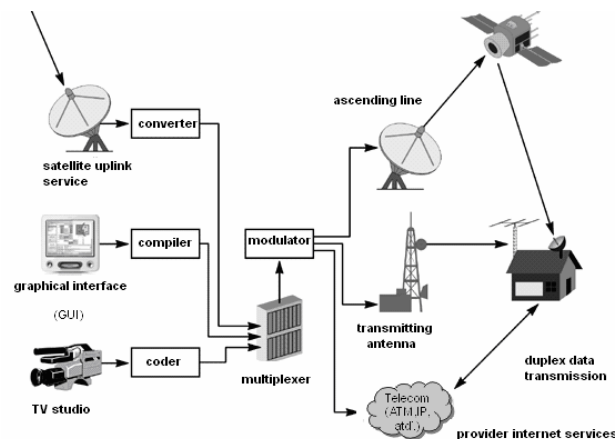


Fig.1 Transportation of contents from studio or downlink channel onto the viewer's screen

II. MHP (MULTIMEDIA HOME PLATFORM)

MHP (Multimedia Home Platform) is a standard on which almost all supplementary services of digital television broadcasting are based, especially the most flexible (interactive) multimedia application (Java language). It was designed by the DVB project. The platform brings technological solutions of customer terminals (receiver DVB-MHP), and enables reception of interactive services. The system provides improved picture, choice of camera scanning a match, interactive access (communication from both sides, user and the other side, server, as well), next the Internet access [4].

Architecture of MHP is composed of three layers: source, software setout and applications. Hardware is not defined. The sources are processing of MPEG format, processor, memory and graphic system. Applications designed for MHP must satisfy the defined application interface API (Application Programming Interface), which is defined by this system [5].

A. Basic layers of MHP

Sources (hardware and software means) needed for picture and sound data processing, e.g. MPEG-2 decoder, or main and graphic processors, which enable access to available application interfaces.

System software – operation software, which contains an application administrator and creates support for basic transportation protocols. Part of it is Java Virtual Machine which separates hardware and software designed by producer from the application interface (API) defined by the norm.

Applications which have access to the platform only through the defined application interface. [5].

The basic layers of MHP are more specified in Fig. 2

According to the Figure 2, an important part is the application manager, with the help of which individual applications are started, managed and ended. For software update, the layer contains so called 'Loader', which is implemented according to a producer.

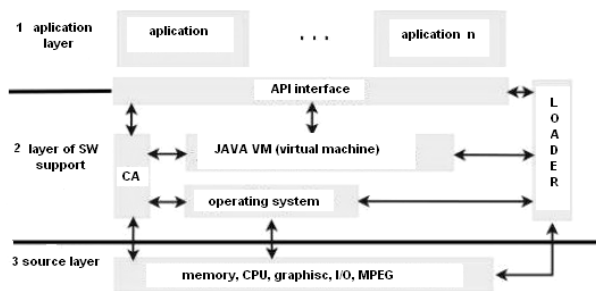


Fig.2. Basic layers and more detailed description of the MHP structure [6],[8].

An important part of the layer is JAVA (JAVA Virtual Machine), which is de facto hardware enclosure for an application layer usage. It means that the defined applications have access to the hardware sources administered through JAVA VM.

In the application layer fall individual programs which are administered by the application manager of the second layer. The applications are started above the level of application interface API (application programming interface), which has strictly defined regulations of access, and therefore enables to separate the application itself from the lower layers and thus reaches multiplatformness [9].

B. Profiles – usage of MHP services)

It is possible to divide MHP into three basic profiles.

Extended television broadcast

Relates to set-top boxes without reverse channel with passive interactivity. Fully defined is in the specification of MHP 1.0.

Interactive television

Also defined in MHP 1.0, but set-top boxes include a reverse channel enabling a higher degree of interactivity. includes extended applications DVB-J API and interactive transportation protocols IP.

The Internet access

set-top boxes with high computing operation, big memory, and the highest degree of interactivity. The profile is defined in MHP 1.1. includes Java API for the Internet access, transportation protocols for broadcasting, DVB-HTML and others.

Currently, the specification MHP 1.2 is being issued, it adds support to provision of DVB services by means of broadband nets (IPTV) and new API. Also, it adds support for providers of the supplied applications, which have been run for the whole period of MHP surrounding. [5].

C. Applications of MHP services

In Figure 3 a block scheme of the transmission chain of digital television signal and MHP is seen. By a provider of multimedia applications we mean a supplier who designs the applications (e.g. weather, television program, crossword puzzles, internet, games, home banking, etc.), and gives it at a

disposal of a user (viewer). Transmission of digital television signal is in progress in three ways. Either it is a satellite transmission (DVB-S), or terrestrial transmission (DVB-T), and finally it is a cable transmission (DVB-C). On the reception side there is a receiver (set-top box), with the help of which user (viewer) chooses the service he is in tune with. Thus, set-top box is the most important element of the transmission chain from the user's point of view, because on it the multimedia applications are started. Transmission of multimedia applications by the MHP standard is based on the program interface API (Application Programming Interface). The programming language Java, which is a multiplatform one, and independent on the used HW and software, was chosen for the multimedia platform.

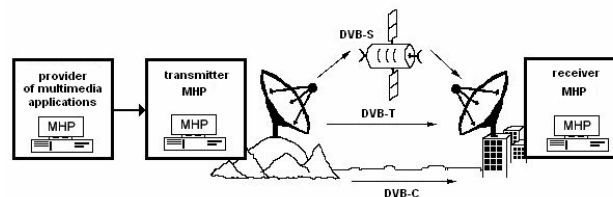


Fig. 3: Transmission and reception of MHP.

The following commercial requirements are posed on the platform: expansibility, possibility of updating, interoperability, the application must be executable for different hardware producers, and for connection providers, separation of data from applications, open standard, system support with conditional access (paid services) (5), open and freely modifiable standard, model of MHP platform in layers, minimum financial duty of initiation of interactive services, Independence of API and CA (Conditional Access System), support of different CA access and Common Interface must be allowed [10].

The applications designed for MHP are accordingly written in JAVA or HTML language. Thus they become independent on hardware platform, or operational system. The MHP platform is being developed by DVB experts. It is designed predominantly for digital transmitters using the DVB-T standard. For really global utilization of MHP, it was necessary to cover the destinations like Japan, USA, where standards of digital transmission different from DVB are used. In consequence, the specification GEM (Globally Executable MHP) has been developed. The specification stems from MHP, but the specifications of DVB standard are eliminated from them. Thereby, compatibility with other standards is allowed [12].

D. Interactive services in MHP

Main subject matter of an interactive television system is ability to execute applications downloaded from the transmitted data stream. One of the basic ideas of MHP is non-existence of an independent MHP application. It means that every MHP application is administered by a service. As viewer uses an MHP application in connection with certain television channel, and switches it over, the application is mostly switched off. This leads to limitation of applications which cannot be run constantly, without being dependent on certain channel (service). They can better guarantee providers that applications will work more properly. This approach also

enables viewer to navigate between applications. Only those which are connected with currently watched program (e.g. voting) will be available. The only transmission pack is insufficient. MHP receiver must ensure which application the given packs belong to, and how to operate them. Accordingly, MHP defines the information application table AIT. It informs receiver about currently available applications. The information is clustered together by components and application manager in MHP receiver. Apart from following available services, the application manager is responsible for monitoring of current services which can execute or end up applications.

MHP applications can be written either in Java or HTML. Apart from this, the MHP platform defines quantity of extensions and limitations in capability of both languages. Therefore, the MHP applications are not compatible with Java or HTML standards. To distinguish, the MHP Java version of the application is designated as DVB-J and HTML version as DVB-HTML [13], [14].

E. HTML applications

Adding of HTML support into MHP, access to internet pages will be enabled. Most of HTML pages are designed for resolution higher than 1024x768, but the resolution of common television is 720x576 only, which will cause general lack of overview.

Application Information Table (AIT). As mentioned above, AIT informs about available applications. But this is not the only provided information, the operating codes follow: AUTOSTART, PRESENT, DESTROY, KILL, PREFETCH, REMOTE.

The application which contains the code AUTOSTART, will be executed automatically after being downloaded into receiver. This allows to execute the applications which are directly related to the currently broadcasted program.

If receiver receives applications with the operating code PRESENT, it will save it in the list of available applications, and user can eventually start them.

Applications with the operating code KILL or DESTROY are shut down by receiver. This is possible to utilize at shutting down certain program, which was the application bound to.

The difference between the codes KILL and DESTROY is that with KILL user has a possibility to continue with the application. With the code DESTROY, application is always shut.

The code REMOTE orders the receiver to go over to certain service in order to execute an application provided by the service.

Next important information states which MHP version is an application assigned (for). This gives a possibility to a player to decide when the application is able to interpret properly, and thus whether it will be enlisted into the list of available services or not (15), (16).

III. TRANSMISSION OF MHP APPLICATIONS

Today, an object carousel is the most frequent method of DSM-CC application transmission. Digital Storage Media – Command and Control (DSM-CC), resp. Object Carousel is a transportation protocol (file system), by means of which it is possible to transfer applications between distributor and receiver. The transfer utilizes encoding of applications into the main transmission stream together with video stream (MPEG2), audio stream data (MPEG1 layer II) and operational data (see pict.4) The data stream is composed of two main parts. The first part comprises the application data themselves (individual files). The second part transfers the Application Information Table AIT, which contains information assigning the given file to a concrete application. Moreover, AIT can contain operating codes: AUTOSTART, PRESENT, DESTROY, KILL, PREFETCH, REMOTE. The codes were defined above.

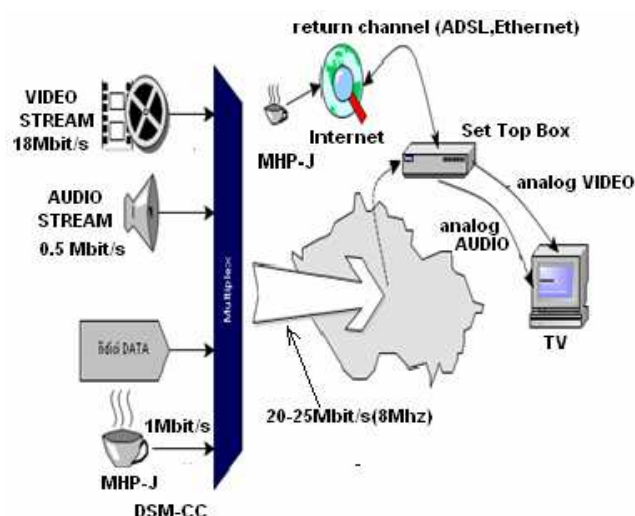


Fig.4 Transmission chain of DVB-T and DSM-CC.

An object carousel is certain package of applications transmitted in a constant loop circuit. This way of transmission brings a problem of a data amount transmitted in the carousel, the amount is directly proportional with the length of the period of one loop, during which the whole process of data transmission is repeated. Length of the period defines what is the longest time receiver has to wait until it is able to start downloading of actual data. In practice, a compromise which will wait acceptable time necessary to download the whole carousel is used.

A. JavaTV

Applications assigned for MHP most often programmed in adapted version of Java language are signed as DVB-J, and adapted Java is designated as JavaTV. The main difference from the standard Java is implementation of the package javax.tv and its version. Here can be found the basic category of the operating run of DVB-J applications. The applications are naturally produced by categories, which are not comprised in javax.tv. For example, reverse channel is programmed exclusively by categories from java.net.

B. Xlet

With an application designed in standard Java, it is assumed that in real time one application will be transferred only. However, digital television needs to execute more applications at the same time. Applet is executed from a web browser, and introduces so called life cycle. This allows to stop applications (when user shifts over from the given web page), and shut it down (when website with applet is shut). At the same time several applets can be executed simultaneously. Therefore, applet was redesigned for the conditions of digital television, and the result is called Xlet. Hence, Xlet contains life cycle (Pict.5), but compared to this, applet consists of the states: 'read', 'start', 'pause', and 'stop'.

The state 'pause' is usable when more applications are being run. In this case, the only visible application will be the one which we are actually working with. The applications in the background will be transferred into the 'pause' mode, and will release sources (memory) for actual computing [12].

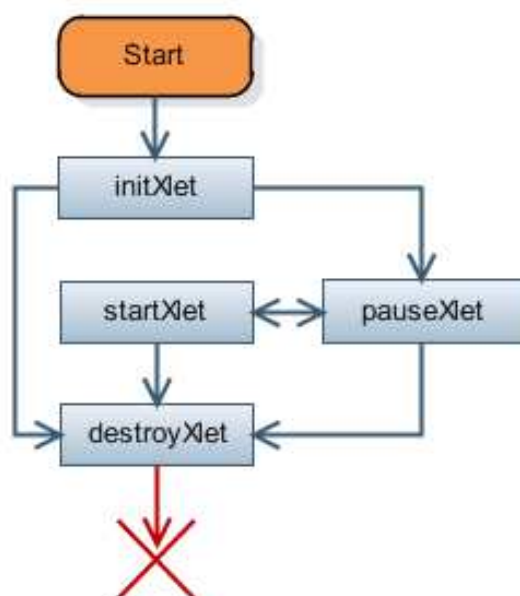


Fig.5: Life cycle of Xlet.

IV. IMPLEMENTATION OF MHP IN PRACTICAL APPLICATIONS

A. JavaTV interface and utilization possibilities

JavaTV is Java ME on a technology which provides higher-performance, and is safe and easy to implement solutions for the Java application development which are run in television, and set-top box settings.

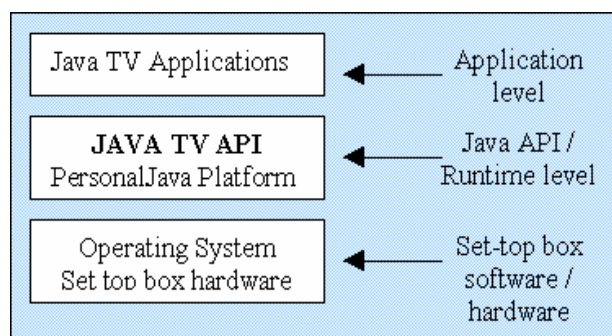


Fig.6 JavaTV

JavaTV offers the developers possibility to design applications like electronic program guide (EPG), Video on Demand (VOD) of clients, games and educational applications, applications for the access to the Internet data (e.g. weather, news, social networks), and most of Blue-ray titles, user interface, bonus contents, see Fig.6.

B. API References – programming core of the JavaTV system

javax.tv.graphics Adds support to AWT with specific television issues, such as mixing in graphics of video, and discovery of the main container with application.

javax.tv.locator provides means of referencing to data and sources available through Java API TV.

javax.tv.media provides operation and actions for administration in real time of media in television surrounding.

javax.tv.net provides mechanism of the access to IP datagrams in transmitted data stream.

javax.tv.service provides API, introduces SI of elements.

javax.tv.service.guide provides API for support of electronic program guide (EPG), including program plans, program of events, and program reviews.

javax.tv.service.navigation provides API for service transfer and hierarchic service information.

javax.tv.service.selection provides mechanism for presentation of service selection.

javax.tv.service.transport provides further information on transportation mechanisms which provide contents of SI data.

javax.tv.util provides interface API for design and administration of timer.

javax.tv.xlet provides interface used by applications, and Application administrator for communication [14].

C. Practical usage of JavaTV

It allows different interactivity of access to EPG, Video on Demand, internet connection, monitoring Facebook, Skype. In Samsung HDTV receivers based on App Store, Fig.7, it allows to start 3 applications of Video on Demand. [17].



Fig.7 Samsung App [17]

Application SPS TV realizes different products of interactivity realized in LCD TV according to Fig.8. For example, Blue-ray players, MP3 players, digital video cameras, and other possible activities on TV.



Fig.8 Multimedia application SPS TV [17].

V. CONCLUSION

In the article, possibilities of IPTV technology utilization today were introduced. In the article, an overview of possible services provided by IPTV technology was introduced. At the same time, the MHP service is described in detail. It is characterised from the point of transfer between receiver and transmitter. The article also describes different layers of the MHP technology, possibilities of its usage in LCD TV. Also it informs about requirements posed on MHP service. The need of knowledge is evident for the reason of recent great extension of the technologies. The MHP application is realised with possibilities in appliances like e.g. set-top box with MHP services, and also services on LCD TV sets. Some of the technologies provide easy and optimum access to the functionality of the whole MHP interface. Visualisation and functionality of the MHP application implementation is described in HTML form together with JavaTV technology, its interface and utilization possibilities. For today's period, great implementation of the above mentioned technologies is necessary, and this is possible only with the means of testing on concrete parameters of transfer, such as transfer speed, reliability, losses, and transfer cuts. It is necessary to be concerned about these parameters, thus investigate reliability of transfer, and other necessary knowledge and parameters of effective MHP service.

REFERENCES

- [1] The ITV dictionary : <http://www.itvdictionary.com/itv.html>, 2001-2003
- [2] BEN: What could we have and what has come, Communicate to them. technique 10/2004, p. II
- [3] SCHWALB, E.: iTV Food Chain; Pearson Education, Prentice Hall PTR, 2005
- [4] LEGÍŇ, M. TV technology DVB-T. Prague, technical references BEN, 2006. ISBN 80-7300-204-3
- [5] LIPTÁK, J., Bachelor thesis of digital television systems, Masarikova University, Faculty of Science 2010.
- [6] www.digizone.cz
- [7] CHMIEL P., KOSTKA T., NAVRÁTILOVÁ E.: Television transmission chain
- [8] BKASAL P.: Interactive applications DVB-MHP , thesis, 2008/2009, UNIVERSITY of Technology, Faculty of Electrical Engineering and Communication.
- [9] MORRIS, S., CHAIGNEAU, A. Interactive TV standards, Amsterdam: Elsevier. 2005. 585 s. ISBN 0-240-80666-2
- [10] LEGÍŇ, M. TVtechnology DVB-T, Praha: BEN, 2006. 286 s. ISBN 80-7300-204-3
- [11] ŠOT, F.: MHP application for remote control of actuating unit, Brno University of Technology, Brno 2008
- [12] ŠVANDA, V.: Methods for image enhancement in interactive applications of digital telecasting, Brno University of Technology, Brno 2008
- [13] EVAİN, J.-P. The Multimedia Home Platform. www.dvb.org/documents/white-papers/evain.pdf (October 2006)
- [14] MORRIS, S.: The JavaTV Tutorial. www.interactivetvweb.org/tutorials/javatv/javatv_intro
- [15] MORRIS, S., CHAIGNEAU, A. Interactive TV standards. 2005. ISBN 0-240-80666-2
- [16] MORRIS, S. The MHP Tutorial. <http://www.interactivetvweb.org/tutorial/mhp/index.shtml.htm> (November 2007)
- [17] <http://www.engadget.com/2010/08/11/samsung-puts-up-cash-hdtvs-for-developers-in-free-the-tv-apps/>

Novel soft switching DC/DC converter

¹Marcel BODOR (3rd year), ²Milan LACKO
Supervisor: ³Jaroslav DUDRIK

^{1,2,3}Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

¹marcel.bodor@tuke.sk, ²milan.lacko@tuke.sk, ³jaroslav.dudrik@tuke.sk

Abstract—A novel full-bridge PWM DC/DC converter with controlled secondary side rectifier using secondary snubber is presented in this paper. Circulating current of the converter was eliminated and soft switching for all power switches of the inverter is achieved for full load range from no-load to short circuit by using controlled rectifier and snubber on the secondary side. The principle of operation is explained and analyzed and the experimental results on a 4,5kW, 100 kHz laboratory model of the converter are presented.

Keywords—Snubber circuit, Soft switching, Zero voltage zero current switching (ZVZCS), DC/DC converter, Controlled output rectifier.

I. INTRODUCTION

PWM high frequency inverter is the conventional way how to control the value of the output quantities in the PWM DC/DC converters. Secondary side of the power transformer is usually fitted by a non controlled centre tapped or bridge rectifier [1] – [4]. The second possible topology of DC/DC converters consists of inverter and controlled output rectifier [5], [6]. A lot of active or passive snubbers, auxiliary circuits or clamps were used in the converters to solve the problem with circulating current and elimination of switching losses [1] – [3]. PWM modulation is often used for the control of the converters. The topology of PWM DC/DC converter which permits all switching devices to operate under soft switching by using controlled rectifier and turn off snubber is described

in the paper.

II. POWER CIRCUIT DESCRIPTION

To improve the properties of the existing converters, new topology of the following DC/DC converter was proposed.

The proposed DC/DC converter shown in Fig. 1 consists of high-frequency full bridge inverter, centre tapped power planar transformer, controlled output rectifier, output filter and novel type of secondary snubber.

The converter is controlled by pulse-width modulation of output rectifier (see Fig. 2). Zero-voltage turn-on and zero-current turn-off all of the transistors T_1 - T_4 in the inverter 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 stored in the leakage inductance of the power transformer 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 in Fig. 3. The DC/DC converter is controlled by modified pulse width modulation with variable phase shift between primary and secondary switches.

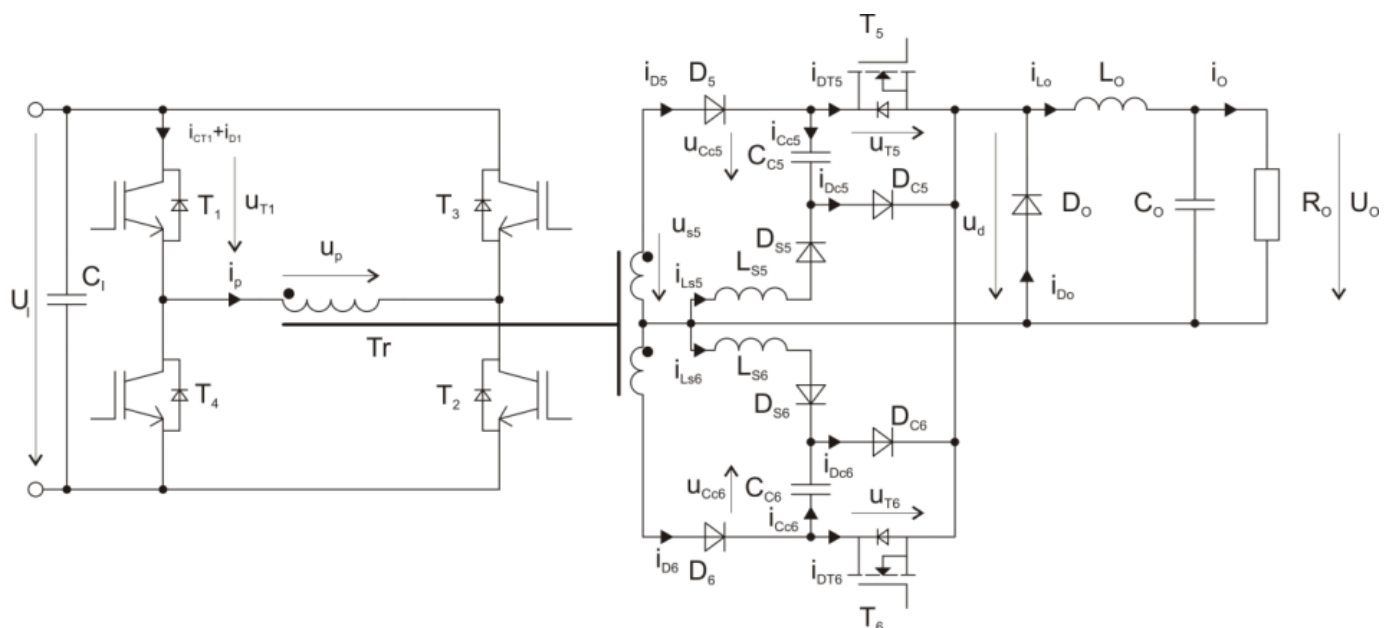


Fig. 1. Scheme of the proposed converter.

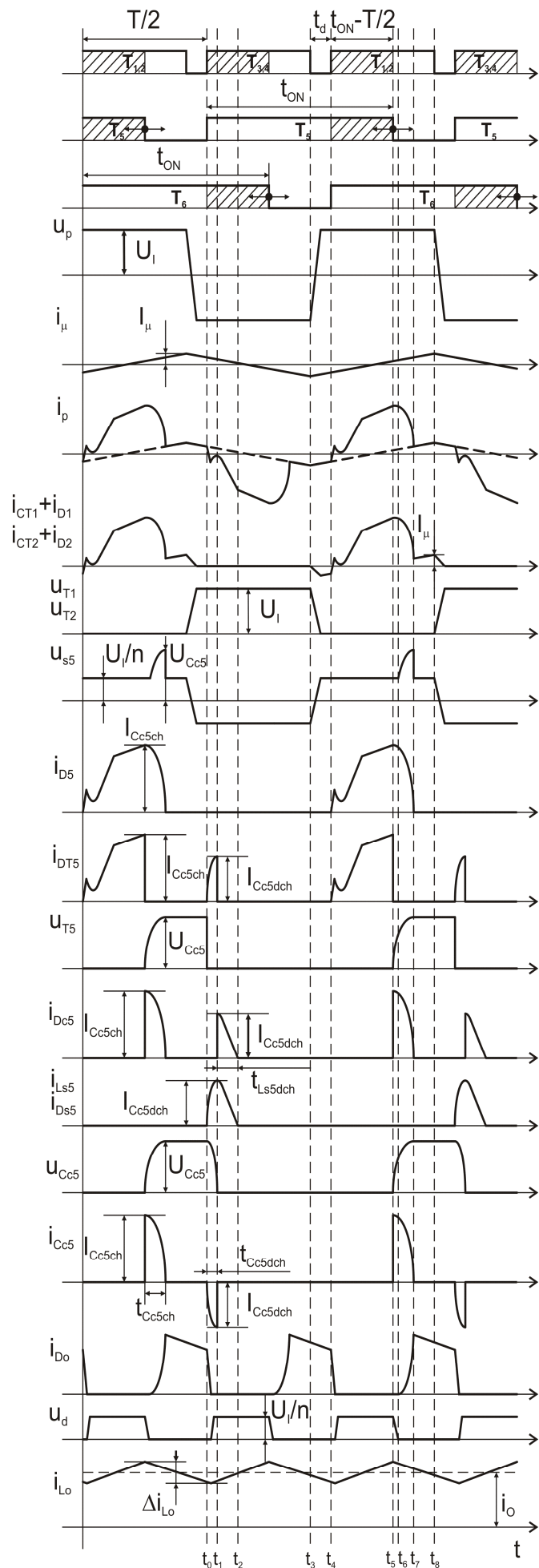


Fig. 2. Operation principle waveforms.

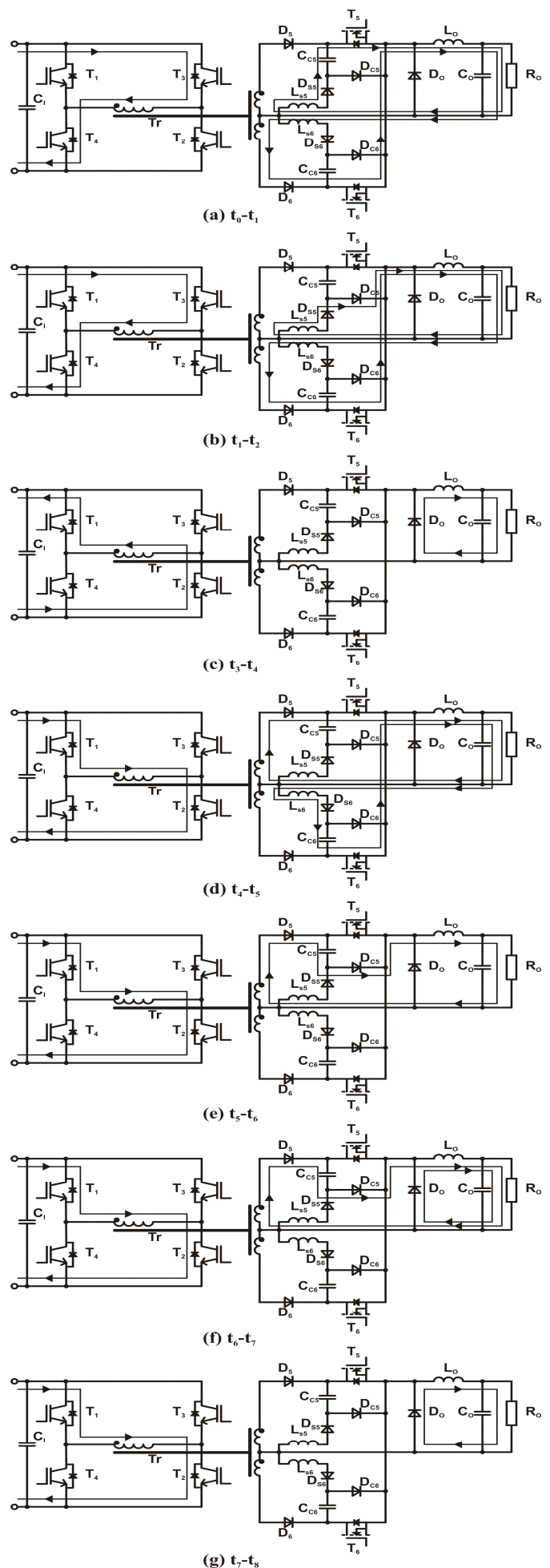


Fig. 3. Operation analysis in the intervals.

Interval (t₀-t₁):

The secondary transistor T₅ is turned on at t₀ half period earlier than primary transistors T₁ and T₂. The capacitor C_{C5} starts discharging through T₅, L_O, R_O, L_S, and D_{S5}. The rate of rise of discharging current of this capacitor C_{C5} is limited by the snubber circuit inductance L_S, and thus zero current turn on for the MOSFET transistor T₅ is achieved. In the same time transistors T₃, T₄ are turned on. Because the transistor T₆ is already in on-state, so the output voltage of the rectifier u_p is equal U_I/n (where n=u_p/u_s is transformer turn's ratio). The current of the primary transistors T₃, T₄ and the current of the secondary transistor T₆ are reduced by the discharging current of capacitor C_{C5}. The equations, which express the time of discharging and amplitude of discharging current, are as follows:

$$t_{Cc5dch} = t_1 - t_0 = \sqrt{L_{S5} \cdot C_{C5}} \cdot \arccos \left(\frac{\frac{U_I}{n}}{\frac{U_I}{n} - U_{Cc5}} \right) \quad (1)$$

$$I_{Cc5dch} = \sqrt{\frac{C_{C5}}{L_{S5}}} \cdot \left(\frac{U_I}{n} - U_{Cc5} \right) \quad (2)$$

Interval (t₁-t₂):

The energy stored in snubber inductance L_S is now flowing through D_{S5}, D_{C5}, L_O, R_O, L_S. At t₂ the whole load current flows through the transistor T₆.

Interval t₃-t₄:

This interval starts with the turn off of the primary transistors T₃ and T₄. The magnetizing current of the transformer Tr discharges the output capacitances C_{OSS} of the transistors T₁, T₂ and charges the output capacitances of the transistors T₃, T₄.

Interval t₄-t₅:

The turn on of the transistors T₁, T₂ and T₆ commutation from freewheeling diode to T₅ transistor occur at t₄. The current of transistor T₅ is reduced by the discharge current of the capacitor C_{C6} and later by the current of the inductance L_{S6}.

Interval t₅-t₆:

At t₅ transistor T₅ turns off. Its current commutates on capacitor C_{C5} and diode D_{C5} and consequently zero voltage turn off of this transistor is ensured. The energy of the leakage inductance of the power transformer is absorbed by the snubber capacitance and then transferred to the load.

Interval t₆-t₇:

At t₆ 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 whole energy of the leakage inductance is absorbed by the capacitor C_{C5}.

At t₇ the current of the rectifier diode D₅ falls to zero, and the primary current that flows through the transistor T₁ and T₂ drops on value of the magnetizing current.

Charging time of the capacitor is:

$$t_{Cc5ch} \approx \frac{\pi}{2} \sqrt{C_{C5} (L'_{1\sigma} + L_{2\sigma})} \quad (3)$$

Voltage of the capacitor after charging is:

$$U_{Cc5} \approx \frac{U_I}{n} + \left(i_o + \frac{\Delta i_{Lo}}{2} \right) \sqrt{\frac{(L'_{1\sigma} + L_{2\sigma})}{C_{C5}}} \quad (4)$$

Interval t₇-t₈:

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 now.

IV. EXPERIMENTAL RESULTS

Laboratory model with components shown in Table I and Table II was built to verify the operation principle of the converter. The converter was supplied from DC source with a value of 325V. The rated output power was 4,5kW at switching frequency of 100 kHz. The typical converter waveforms were obtained at nominal output voltage of 45V and output current of 90A.

TABLE I
USED COMPONENTS IN CONVERTER

T ₁ -T ₄	IRG4PSC71UD
T ₅ ,T ₆	2xIRFP4568
D ₅ ,D ₆ D _{C5} , D _{S5} , D _{C6} , D _{S6}	150EBU02
D _O	2x150EBU02

TABLE II
USED PLANAR TRANSFORMER PARAMETERS

Core	2xE64
Material	3C90
Turns ratio	1:1.5
R ₁	6,5mΩ
R ₂	0,75mΩ
L _{1H}	706μH
L _{1σ}	310nH
L _{2σ}	12,6nH

Primary transistor collector-emitter voltage and collector current with gate signals of primary and secondary transistors are shown in Fig. 4. After drop of the secondary MOSFET transistor gate signal u_{GST5} to zero, the transformer primary current sinks 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. At the turn on moment of primary IGBT transistors the primary current (transformer magnetizing current) discharges their output capacitance 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. It is evident that operating point of the transistor is moving in the low losses area.

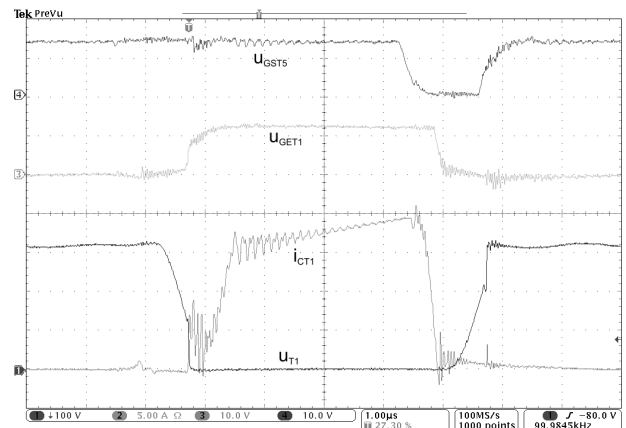


Fig. 4. Primary transistor voltage and current at turn on and turn off.

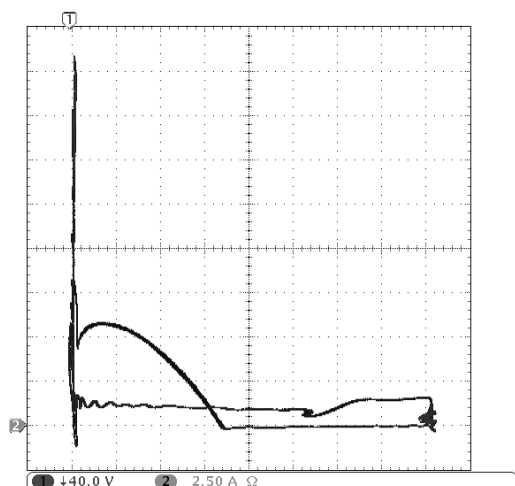


Fig. 5. Switching trajectory of the primary transistor.

Primary voltage and current of power planar transformer are shown in Fig. 6. The switches (T_1 , T_4) and (T_3 , T_4) are operated very simple as two pairs with duty cycle of 0.5 and therefore the primary voltage is rectangular. It can be seen that circulating current is totally suppressed and only magnetizing current of the transformer flows during freewheeling interval.

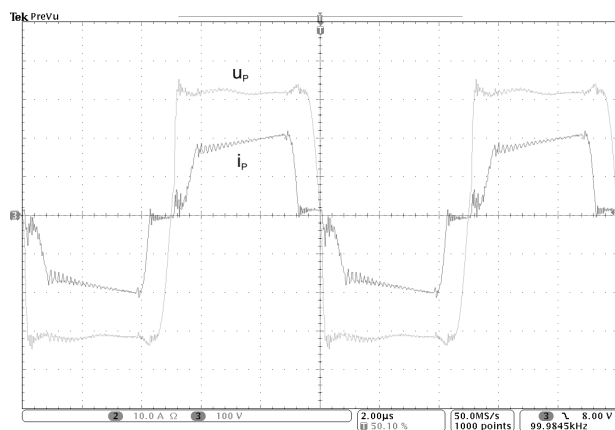


Fig. 6. Primary voltage and current of the transformer.

Secondary transistor drain-source voltage and drain current at turn on and turn off are shown in Fig. 7. At turn off of the MOSFET transistor its collector current commutates on the snubber capacitor 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} .

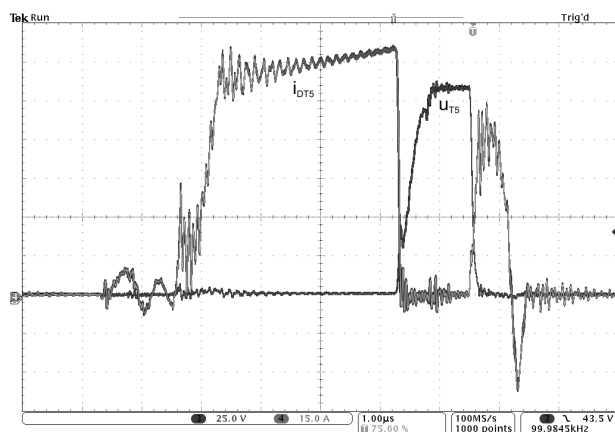


Fig. 7. Secondary transistor voltage and current at turn on and turn off.

Switching trajectory of secondary transistor is shown in Fig. 8.

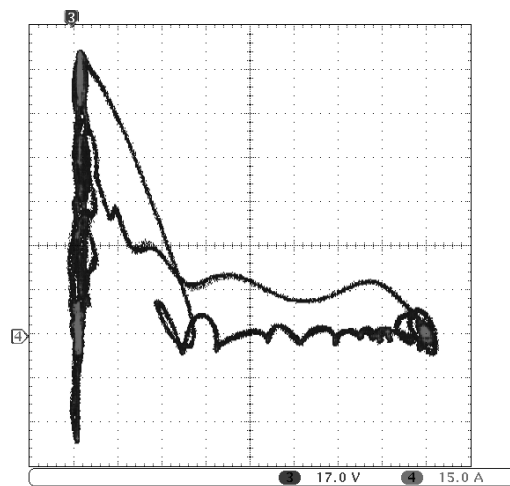


Fig. 8. Switching trajectory of the secondary transistor.

Efficiency of the converter at output voltage of 45V is shown in Fig. 9. Efficiency of the converter is around 94%.

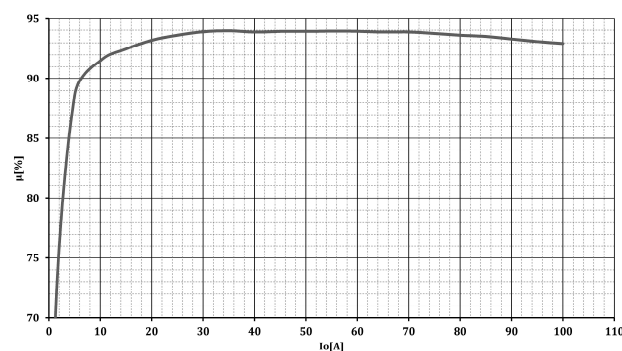


Fig. 9. Efficiency of the converter.

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.

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.

ACKNOWLEDGEMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0185-10.

REFERENCES

- [1] J. Dudrik, *High Frequency Soft Switching DC-DC Power Converters*, Monograph, Elfa, Košice, Slovakia, 2007 (in Slovak).
- [2] P. Chlebiš, *Soft Switching Converters*, Monograph, VŠB-TU Ostrava, Ostrava, 2004, ISBN 80-248-0643-6, 148 p. (in Czech)
- [3] 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.
- [4] S. J. Jeon, G. H. Cho, A Zero-Voltage and Zero Current Switching Full Bridge DC-DC Converter With Transformer Isolation, *IEEE Trans. on Power Electronics*, Vol.16, No.5, 2001, pp. 573 – 580.
- [5] J. Dudrik, N. D. Trip, Soft- Switching PS-PWM DC-DC Converter for Full-load Range Applications, *IEEE Transactions on Industrial Electronics*, Vol. 57, no. 8 (2010), p. 2807-2814. ISSN 0278-0046
- [6] 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.

Overview of the selected technologies for creating the virtual laboratory for the assembling technologies in electronics

¹Dominik DEMETER (2nd year)
Supervisor: ²Juraj BANSKÝ

^{1,2}Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic

¹dominik.demeter@tuke.sk, ²juraj.bansky@tuke.sk

Abstract — Creating a virtual laboratory is a complex task which includes the implementation of different technologies. Different technologies are needed for creating the user interface, the education material and for simulating the experiments. This paper is an overview of the technologies which will be used for creating the virtual laboratory at the Department of Technologies in Electronics which will be focused on the e-Learning education of assembly technologies in electronics.

Keywords — e-Learning, digital video, remote desktop, virtual laboratory, virtual tour.

I. INTRODUCTION

The Department of Technologies in Electronics has several reasons which led us to a decision to create a virtual laboratory. The main reason is that the Department has started a new study program called Automotive Electronics which increased the number of students. This new situation has remitted to other problems, like bad space disposition of the laboratories and causes increased financial load for laboratory experiments and higher load on teachers. The increased number of students is also a potential threat for the laboratory equipment which are very sensitive for the cleanness of the laboratories.

Our goal is to create a virtual laboratory which will reduce these negative impacts of the increased number of students and will help us to offer the students at least the same quality of education, as before these changes. After the implementation of the virtual laboratory our attention will be focused on increasing the quality of the education by including the education of some new technologies which are not physically available at our Department.

II. TECHNOLOGIES FOR CREATING THE VIRTUAL LABORATORY

A. User interface

The user interface of the virtual laboratory is a very important part of the complex e-Learning system because this gives the first impact for the students and affects the whole success of the education material. The user interface must be easy to use and must be available for the users immediately, without any complicated installation procedure. Our virtual

laboratory will be based on standard technologies which are available in every modern operation system. We will use for implementing the user interface of the virtual laboratory technologies HTML, PHP, Flash, JAVA and JavaScript.

At our department, we are trying to give to the students the feeling as they were physically present at the laboratories. For this reason, from every laboratory we will create a virtual tour. The virtual tours will be created from one or more 360° panorama pictures (Fig. 2).

The panorama pictures will be integrated to a HTML page by an open-source JAVA applet, called PTVIEWER [1]. This applet allows us to show the panorama pictures as virtual tours with standard navigation buttons and instructions for automatic turning around, and others. The most important feature of this applet for the virtual laboratory is the ability to create some active areas which can be used as hypertext links for opening new web pages or trigger some JavaScript (like image maps in HTML pages). These links are used for interconnecting some different panorama pictures to one virtual tour, and are used for marking the laboratory equipment. The active areas are highlighted (Fig. 1), when the



Fig. 1. Demonstration of the Virtual tour with selected (highlighted) technological equipment.

pointer gets over these areas. There is also available a button to highlight every available active area to help the users navigation.



Fig. 2. Example of the 360° panorama picture, which is used for creating the Virtual tour of the laboratory.

Integrated part of the virtual laboratory will be a database of the available laboratory equipment. This database will contain useful information about the laboratory equipment, such as: technical description (including the name, type, manufacturer name); user manuals; the name of the laboratory in which the equipment is located in, including a hypertext link to the virtual tour of the appropriate laboratory; detailed photography about the equipment; list of technological experiment which the equipment can be used for; link to the education materials, in which the selected laboratory equipment can be found; safety warning and instructions; etc.

B. Education materials

The virtual laboratory can meet our needs only in case, if it will offer the students some education materials which can compensate the classical laboratory experiments. For this reason, we have decided to create from every laboratory experiment an edited video clip with audio track where the speaker will explain every step of the experiment.

Digital video is a very important building brick of a good e-Learning material because digital video can shows the students the same laboratory experiment several times from several points of view. A well prepared video clip can contains some inserted parts, like some animations, static pictures. The author of the education material can also work with the timeline of the video clip, and slow down the video to focus the attention of the students on to some selected part of the experiment. The video can be also fasted up, in case the laboratory experiment is time consuming and show only the interesting parts.

The biggest advantage compared to standard education form with practical experiment exhibitions is, that the students can watch the same video many times and catch every detail of the experiment. The experiments are available for every student, any time which brings more flexibility to the education process, and can help the study of the students, who cannot be present at the classical laboratory experiments (e.g. handicapped students). This form of education also allows us to demonstrate laboratory experiments and technologies which are not physically available at our laboratories.

Our virtual laboratory will contain a database of video clips which will be interconnected with the laboratory equipment

which are available in our laboratories. We are also planning a database of available laboratory experiments which will be linked to the video database.

The database of the prepared video files will not be the only way, how digital video will be represented in the virtual laboratory. We are planning to integrate to the virtual laboratory some live video streams from the laboratories. This video streams will be used for showing some real time laboratory experiments. For this needs, we have available three IP cameras which can be directly connected to the computer network. The IP cameras have connectors for audio input and output which allows two way communications between the students and the laboratory stuff. This can be useful in case the live video stream is showed in the classroom and the teacher is exploring the laboratory experiments for students. These IP cameras can be also used for controlling the laboratory equipment and experiments during executing the remote laboratory experiments.

We have also available a digital camcorder which will be used for creating the video clips from the laboratory experiments. This camera can be used for live video streaming, too. For this needs, we bought a TV card which can be used to get the live picture from the camcorder to a desktop PC or notebook and to stream through the computer network.

The integrated part of the virtual laboratory will be also digital video clips created from the desktop of the teacher's computer, supplemented with the audio track created during the lecture. This will help the students for preparing for their final exam by watching the lectures more times.

C. Virtual experiments

The most complicated, but very useful part of a virtual laboratory is a user interface which gives the students the possibility to executing virtual laboratory experiments. We can recognize two types of virtual laboratory experiments: fully virtual experiments which are available for more students in the same time because it is in fact only a computer program; remote experiments on real laboratory equipment which are available only for one user at time because it is technically impossible to do different thing at the same equipment.

We are planning to integrate to the virtual laboratory some remote experiments on real laboratory equipment. Only a few

numbers of equipment have built-in Ethernet interface. For this reason, we need to find the way, how to connect the laboratory equipment to the computer network, and how communicate with them remotely. Laboratory equipment has at least a serial communication interface which can be used for communication with a computer. Most of the laboratory equipment is connected directly to a dedicated computer.

For creating remote access to the laboratory and to the laboratory equipment, the simplest way is to create remote desktop connection to the connected computers and take control over the laboratory equipment. This is the best solution because the students will have also available the software which is installed on the connected computer and which is sometimes bundled with the laboratory equipment and cannot be installed on the computers where the laboratory equipment's connection is not detected. This solution is preferred because the computers are still available and it can be realized without high financial investments. The negative feature of this solution is the fact, that the users have full access to the laboratory equipment and we cannot take control over the way of using the equipment. We also need to create in the operation system of the connected computers a user profile which will have restricted access to the settings of the computer to guarantee an error-less working and which can be easily replaced after every reservation ended, in case of any problem occurs. Because the virtual laboratory has a Web based interface, we need to find a way, how to integrate a Remote Desktop connection into a web page. The most appropriate open-source software solution which will be tested to implement a Remote Desktop connection to the virtual laboratory's web interface is ThinVNC. [2]

Other possible solution is to use some Serial to Ethernet convertors, such as Lantronix's XPort [3] in combination with their software solution called Com Port Redirector [4]. This allows us to use the equipment with serial connection over the computer network because the software creates a virtual communication port on the user's computer (Fig. 3). This solution is available only for those laboratory equipment, where we do not need to have installed some licensed software.

The most complicated way, how to create a remote access is a combination of a Serial to Ethernet convertor with our own

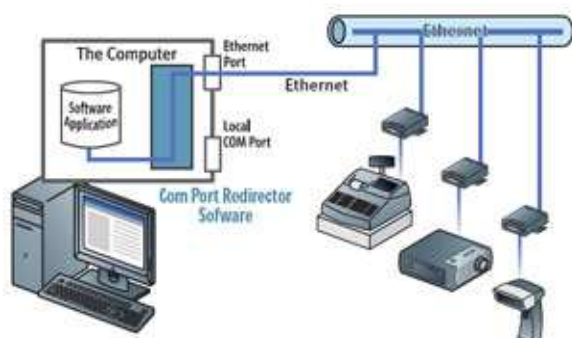


Fig. 3. Example of use the Lantronix XPort for remote controlling the laboratory equipment through Ethernet network.

user interface. This way of connection seems to be the best solution because we can take control over every users command and the way of using the equipment, but it is very hard to create a complex user interface similar to the bundled software. This solution is usable mostly for prepared laboratory experiment, e.g. some measuring where we need to

get into the virtual laboratory only some data, and we have just a few options, how to affect the final result of the experiment. In case of using this solution, it is necessary to have laboratory equipment which have implemented some open-source communication protocol or have available a technical documentation with detailed description of the used communication protocols and the available commands.

The experiments based on remote control needs an accurate reservation system with implemented access control mechanism to guarantee the students an exclusive access to the computers and equipment in the selected time, and for schedule the access of the high number of students.

D. Background technologies

Today's most used communication channel for e-Learning solutions are local computer networks and internet. Our virtual laboratory will be connected to our LAN and some features will be accessible through internet. This needs to have the LAN properly prepared for the enhanced load.

We need to create some LAN connection ports to every laboratory to be able to connect the computers, laboratory equipment and IP cameras to the LAN and internet. We also need to replace the old 10 Mbit/s fast hubs with new 100 Mbit/s switches with gigabit uplink to the server. The computer network with these parameters will be suitable for parallel streaming from our IP cameras without limiting the other users of the LAN.

III. CONCLUSION

The described virtual laboratory is a complex e-Learning solution which will help our Department to handle the increased number of students as well as it will offer a new education material for our distant students. We are working on the creation of the laboratory continuously, some features are still implemented and some of them are under creation. The implementation of the whole virtual laboratory will be finished during the next year.

ACKNOWLEDGEMENT



This paper was developed with support of the project "Centrum excelentnosti integrovaného výskumu a využitia progresívnych materiálov a technológií v oblasti automobilovej elektroniky" (Centre of Excellence of the Integrated Research & Exploitation the Advanced Materials and Technologies in the Automotive Electronics"), ITMS 26220120055, that is co-financed from Structural Funds EU ERDF within Operational Program Research and Development OPVaV-2009/2.1/03-SORO and preferred axis 2 Support of Research and Development.

REFERENCES

- [1] The documentation of the PTViewer, <http://webuser.fh-furtwangen.de/~dersch/PTVJ/doc.html>
- [2] ThinVNC - HTML5 Remote Desktop, <http://sourceforge.net/projects/thinvnc/>
- [3] XPort, Embedded Ethernet Device Server, <http://www.lantronix.com/device-networking/embedded-device-servers/xport.html>
- [4] Com Port Redirector, <http://www.lantronix.com/device-networking/utilities-tools/com-port-redirector.html>

Partial discharges pattern statistical analysis

¹Marián HRINKO (2nd year), ²Marek PAVLIK (1st year)
Supervisor: ³Iraida KOLCUNOVÁ

¹Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic

^{2,3}Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic

¹marian.hrinko@tuke.sk, ²marek.pavlik@tuke.sk, ³iraida.kolcunova@tuke.sk

Abstract — This article deals about statistical processing of measured data from partial discharge measurements. From measured data are calculated descriptive operators as skewness, kurtosis, charge asymmetry and cross correlation factor that could describe partial discharges patterns by means of one number. These numbers could be useful for recognition of partial discharges patterns by means of neural network.

Keywords — Partial discharges, statistical operators, skewness, kurtosis, cross-correlation.

I. INTRODUCTION

Discharge measurement is an important means of testing the reliability of insulation system of HV rotating machines. There are many successful methods for detection, location and evaluation partial discharge phenomena. These quantities do not predict the lifetime of insulation systems but give information about thier quality [1].

As each defect has its own particular degradation mechanism, it is important to know the correlation between discharge patterns and the kind of defect [1]. To describe of partial discharge patterns are used descriptive operators that give numerical information about the shape of partial discharge patterns. This information could be sent as input value to neural network to evaluate patterns of partial discharges and so can be created the autonomic classification system of defects in insulation system.

II. DESCRIPTION OF MEASURED OBJECT

To obtain the necessary information for statistical processing (apparent charges and frequency) has been measured five high-voltage coils.

The rated voltage of coil was $U_n = 6\text{kV}$ and the rated power was $P_n = 0,2\text{ MW}$. The main insulation of measured coils was Samicatherm 366.28 – insulation created by mica–calcined tape amalgamated by epoxy asphalt by means of Resin Rich technology. All coils had semi-conductive protection CoronaShiled P800 – a modified phenolic resin with semi-conductive filler.

III. SET UP AND MEASUREMENT

On the detection of partial discharges was used off-line method with direct galvanic coupling according to IEC

60270 (see fig.1), where the measured impedance was connected in to series with coupling capacitor. This set up ensures higher safety of measured impedance in the case of breakdown sample.

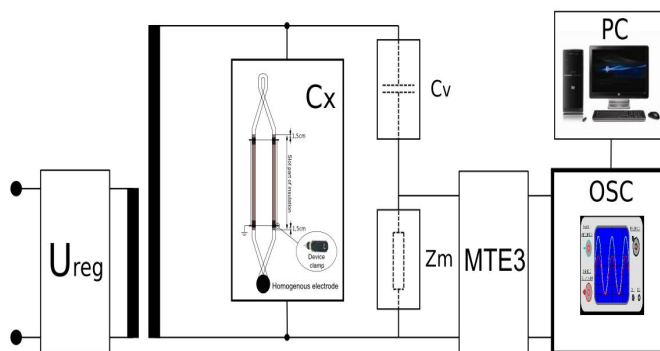


Fig.1 Scheme of experimental setup.

Measurement was done in laboratory conditions. The coil was hung on insulation rope placed in Faraday cage that defended the crossing of interfered electromagnetic waves from surround outside space. The end of coil was galvanic joined by homogenous electrode. The main purpose of homogenous electrode besides of galvanic connection was eliminated the corona discharges rose from harsh protrusions. In order to model of putting coil in the stator slot, the slot part of coil was grounded by means of galvanic joint of four points wrapped by conductive foil.

Measurement was done from voltage 2 kV with step 200 V to voltage 6 kV. The program for evaluating and denoting discharge activity had activated in the moment, when stabile discharge activity was captured. Every measurement took three minutes and during the time were captured 440 (480) periods of voltage cycles.

IV. STATISTICAL ANALYSIS

Measured data were stored in two files. The first file contains values of apparent charges trapped in time moments and in the second file there are stored frequencies of trapped apparent charges. In practice it is the matrix with 440 (480) rows and 199 columns. The particular rows represent the number of captured voltage cycles and the particular columns represent number of phase windows. It follows that one period of applied voltage is divide into 199 phase windows so to one phase window appertain $1,81^\circ$.

From measured data were created histograms of distribution $H_{q_{\max}}(\varphi)$, $H_{q_n}(\varphi)$ and $H_n(\varphi)$ by means of special program.

$H_{q_{\max}}(\varphi)$ represents a vector of maximal values of apparent charges captured in individual phase windows. It can be mathematical described as:

$$H_{q_{\max},i} = \max_{1 \leq j \leq n_i} [q_{i,j}] \quad (1)$$

$H_{q_n}(\varphi)$ represents a vector of mean values of apparent charges captured in individual phase windows. The mean charge value is calculated as a fraction of sum apparent charges in particular phase windows to frequencies in the same phase windows. It can be mathematical described as:

$$H_{q_n,i} = \frac{1}{n_i} \sum_{j=1}^{n_i} q_j n_j \quad (2)$$

$H_n(\varphi)$ represents a vector of frequencies charges in individual phase windows. In practice it is the mean value of particular columns of file with frequencies. It can be mathematical described as:

$$H_{n,i} = \frac{1}{n_i} \sum_{j=1}^{n_i} n_j \quad (3)$$

In practice, partial discharges occur in two sequences during the voltage cycle: in positive half of period and in negative half of period. Therefore the distribution of $H_{q_{\max}}(\varphi)$, $H_{q_n}(\varphi)$ and $H_n(\varphi)$ are characterized for positive and negative half of period too: $H_{q_{\max}}^+(\varphi)$, $H_{q_n}^+(\varphi)$ and $H_n^+(\varphi)$, $H_{q_{\max}}^-(\varphi)$, $H_{q_n}^-(\varphi)$ and $H_n^-(\varphi)$. From distributions of $H_{q_{\max}}^+(\varphi)$, $H_{q_n}^+(\varphi)$ and $H_n^+(\varphi)$, $H_{q_{\max}}^-(\varphi)$, $H_{q_n}^-(\varphi)$ and $H_n^-(\varphi)$ are depicted histograms of distributions.

For the purpose of simplify description of histograms distributions $H_{q_{\max}}^+(\varphi)$, $H_{q_n}^+(\varphi)$ and $H_n^+(\varphi)$, $H_{q_{\max}}^-(\varphi)$, $H_{q_n}^-(\varphi)$ and $H_n^-(\varphi)$ were used descriptive operators as skewness, kurtosis, charge asymmetry and cross correlation factor.

The first descriptive operator is skewness γ_1 , which is defined by means of third moment:

$$\gamma_1 = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{1}{n} \frac{\sum_{i=1}^n (q_i - \bar{q})^3 n_i}{s^3} \quad (4)$$

where q_i are values of the maximal pulse height distribution $H_{q_{\max}}^+(\varphi)$, $H_{q_{\max}}^-(\varphi)$, the mean pulse height distribution $H_{q_n}^+(\varphi)$, $H_{q_n}^-(\varphi)$ and the pulse count distribution $H_n^+(\varphi)$, $H_n^-(\varphi)$ for positive and negative half of voltage cycle;

and \bar{q} is the arithmetic average of the maximal pulse height distribution $H_{q_{\max}}^+(\varphi)$, $H_{q_{\max}}^-(\varphi)$, the mean pulse height distribution $H_{q_n}^+(\varphi)$, $H_{q_n}^-(\varphi)$ and the pulse count distribution $H_n^+(\varphi)$, $H_n^-(\varphi)$ for positive and negative half of voltage cycle.

$$\bar{q} = \frac{1}{n} \sum_{i=1}^n q_i n_i \quad (5)$$

s is a standard deviation from average value of the maximal pulse height distribution $H_{q_{\max}}^+(\varphi)$, $H_{q_{\max}}^-(\varphi)$, the mean pulse height distribution $H_{q_n}^+(\varphi)$, $H_{q_n}^-(\varphi)$ and the pulse count distribution $H_n^+(\varphi)$, $H_n^-(\varphi)$ for positive and negative half of voltage cycle.

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (q_i - \bar{q})^2 n_i} \quad (6)$$

The second descriptive operator is kurtosis γ_2 , which is defined by means of fourth moment:

$$\gamma_2 = \frac{\mu_4}{\sqrt{\mu_2^4}} - 3 = \frac{1}{n} \frac{\sum_{i=1}^n (q_i - \bar{q})^4 n_i}{s^4} - 3 \quad (7)$$

The symbols q_i , \bar{q} and s have the same meaning as in the case of skewness. The normal distribution has kurtosis equal number 3 so for reaching zero value is necessary subtract number 3.

The third descriptive operator is charge asymmetry Q , which is defined as:

$$Q = \frac{Q^- / N^-}{Q^+ / N^+} = \frac{N^+ \sum_{i=n/2}^n q_i}{N^- \sum_{i=1}^{n/2} q_i} \quad (8)$$

where Q^+ and Q^- are sum of charges in positive half of period and in negative half of period and N^+ and N^- are frequencies of apparent charges in positive and in negative half of period.

The fourth descriptive operator is cross correlation factor cc , which is defined as:

$$cc = \sqrt{\frac{[\text{cov}(Q^-, Q^+)]^2}{\text{cov}(Q^-, Q^-) \text{cov}(Q^+, Q^+)}} \quad (9)$$

V. PROCESS OF MEASURED DATA

The main goal of this article was find out the dependence of skewness and kurtosis of the mean pulse height distribution $H_{q_n}(\varphi)$ and the pulse count distribution $H_n(\varphi)$ on applied voltage.

The first important step was processed the measured data into graphs $H_{q_{\max}}(\varphi)$, $H_{q_n}(\varphi)$ and $H_n(\varphi)$ so-called patterns of partial discharges (see Fig. 2, Fig. 3, Fig. 4). Next step was calculated descriptive operators skewness S_k , kurtosis K_u , charge asymmetry Q and cross correlation factor cc from patterns of partial discharges $H_{q_{\max}}(\varphi)$, $H_{q_n}(\varphi)$ and $H_n(\varphi)$.

The comparison of skewness and kurtosis on applied voltage were done only for the mean pulse height distribution $H_{qn}(\varphi)$ and the pulse count distribution $H_n(\varphi)$ cause importance of these two distributions [1].

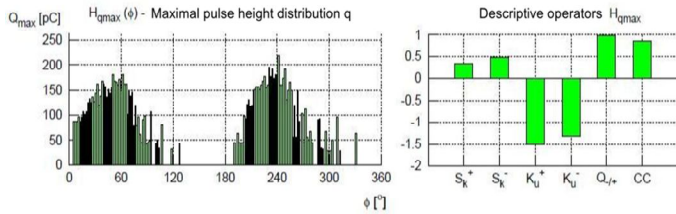


Fig.2. Graphs of the maximal apparent charge distribution $H_{qmax}(\varphi)$, and corresponding descriptive operators for voltage value 3,6 kV.

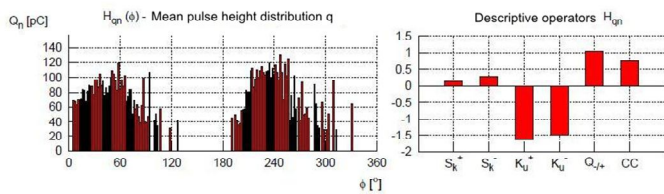


Fig.3. Graphs of the mean apparent charge distribution $H_{qn}(\varphi)$, and corresponding descriptive operators for voltage value 3,6 kV.

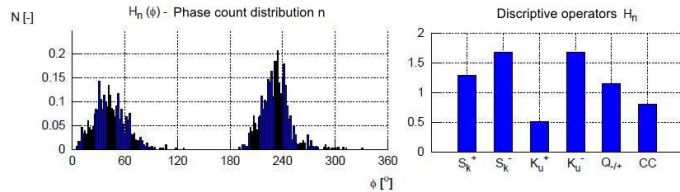


Fig.4. Graphs of the frequencies apparent charge distribution $H_n(\varphi)$, and corresponding descriptive operators for voltage value 3,6 kV.

VI. REVIEWED OF RESULTS

On the Fig. 5 there is depicted dependence of skewness on applied voltage. It can be seen that values of skewness fall down with raising voltage in positive and in negative half of period. It finally means that distribution of the mean apparent charge is slowly turned to the right side.

On the Fig. 6 there is depicted dependence of kurtosis on applied voltage. It can be seen that kurtosis raise up with raising voltage in both half of period. It means that the mean pulse height distribution $H_{qn}(\varphi)$ became pointer with raising voltage.

Fig. 7 shows dependence of skewness on applied voltage. It can be seen that skewness fall down in major part of measured coils with raising voltage. In the case of the pulse count distribution $H_n(\varphi)$ turning to the right side is weaker than in the case of the mean pulse height distribution $H_{qn}(\varphi)$.

According to the dependence of kurtosis on applied voltage on the Fig. 8 it can be seen that kurtosis of the pulse count distribution $H_n(\varphi)$ was changing to lower values with raising voltage. It means that the pulse count distribution became flatter with higher voltage.

In the local place, where are discontinuities of distributions there was changing area of reinforcement of the measured equipment.

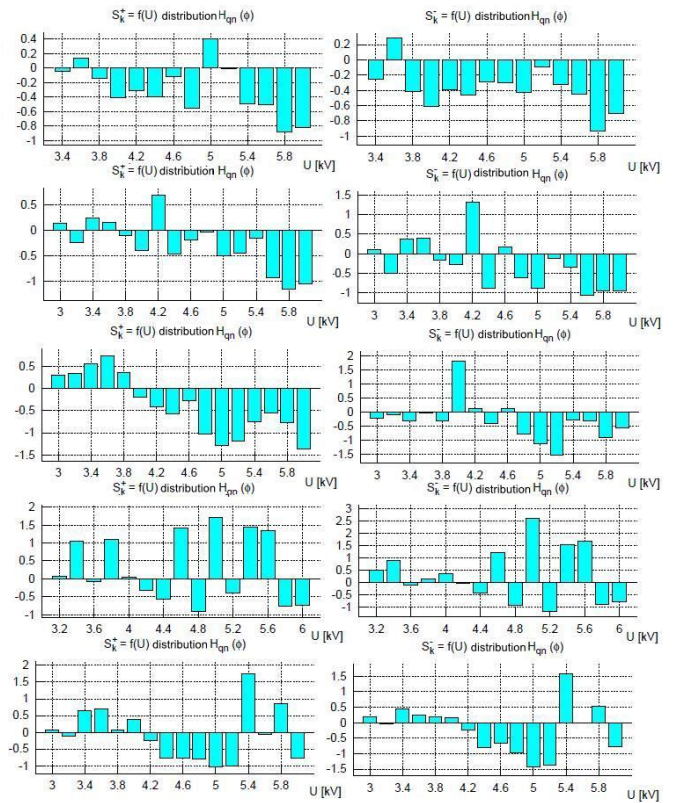


Fig.5. Dependence of skewness of the mean apparent charge $H_{qn}(\varphi)$ on applied voltage.

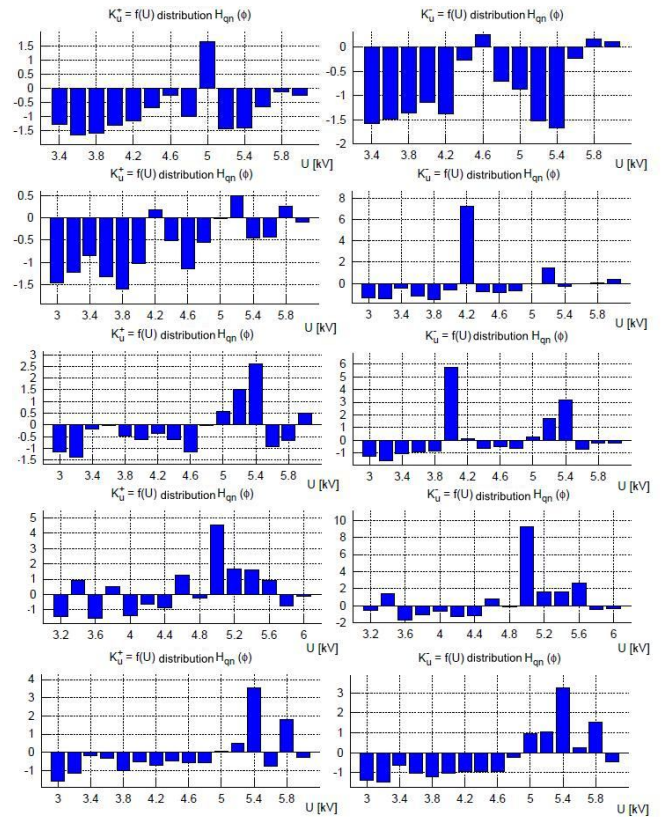
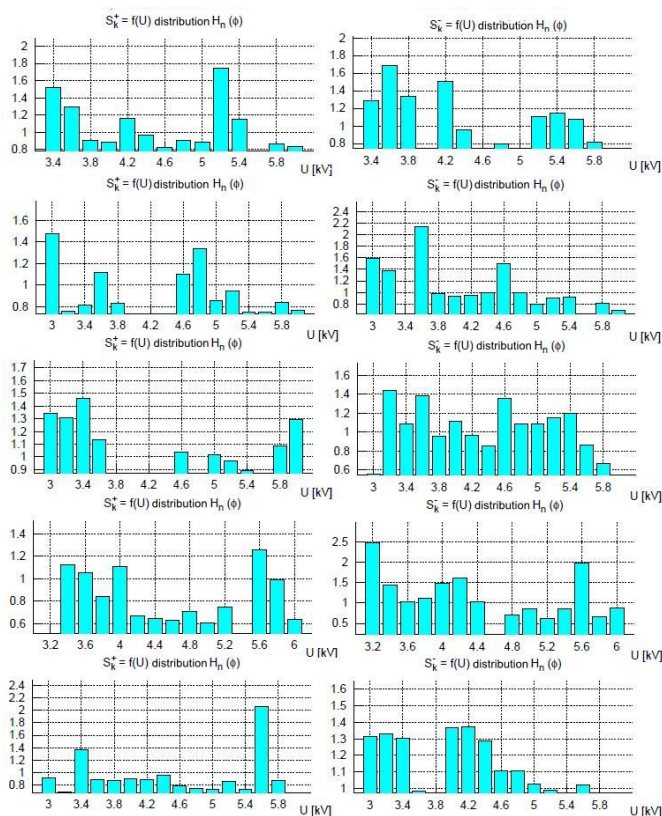


Fig.6. Dependence of skewness of the mean apparent charge $H_{qn}(\varphi)$ on applied voltage.

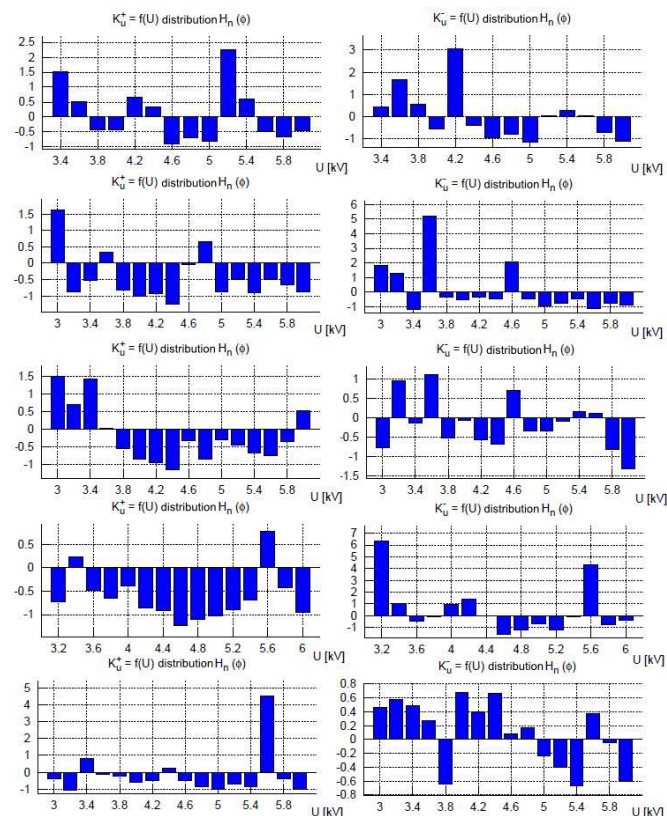
Fig.7. Dependence of skewness of the apparent charge frequencies $H_n(\varphi)$ on applied voltage.

VII. CONCLUSION

In this article was observed dependence of skewness and kurtosis on applied voltage. There was effort find these descriptive operators that the best indicate type of the partial discharges activity.

The best indicator seems to be skewness of the mean pulse height distribution. In the course of lower voltage prevails internal discharge activity and values of skewness reached values near to zero – distribution obliquity was resembling to normal distribution. In higher voltage values, where is expressive surface discharges activity, skewness reached lower values – the distribution turned to right in compare with normal distribution.

These observations could be useful for autonomic partial discharge recognition.

Fig.8. Dependence of kurtosis of the apparent charge frequencies $H_n(\varphi)$ on applied voltage.

REFERENCES

- [1] E. Gulski, F. H. Kreuger, *Computer-aided recognition of Discharge Sources*, IEEE Transactions on Electrical Insulation, Vol. 27 No. 1, February 1992, pp. 82 – 82. ISBN 0018-9367.
- [2] I. Kolcunová, *Diagnostika elektrických strojů*, Technická Univerzita Košice, 2006. ISBN 80-8073-550-6
- [3] N.C. Sahoo, M.M.A.Salama, R. Bartnikas. *Trends in Partial Discharge Pattern Classification: A Survey*. IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 12, No. 2, April 2005, pp. 248 - 264. ISSN 1070-9878.
- [4] T. Tanaka, *Partial discharge pulse distribution pattern analysis*. IEEE Science, Measurement and Technology, Tokyo, Jan 1995, Vol. 142, pp 46 - 50. ISSN 1350-2344.
- [5] E. Gulski, F. H. Kreuger A. Krivda, *Classification of Partial Discharges*, IEEE Transactions on Electrical Insulation, Vol. 27 No. 1, December 1993, pp. 917 – 931. ISSN 0018-9367.

Road Signs Recognition and Inventory System using Optical Correlator

¹František RAKOCI (1st year)
Supervisor: ²Luboš OVSENÍK

^{1,2}Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

¹frantisek.rakoci@gmail.com, ²lubos.ovsenik@tuke.sk

Abstract—This article describes usage of Fourier optical transformation in the road signs recognition process. The main role of proposed system is to evaluate existence of road sign similarity with using of optical correlator. This evaluation is based on prepared database in logged image. Our goal is to propose device for road signs inventory. The device have to be able to log the existence of signs in video log and specify the correct location of signs and their conditions.

Keywords—fourier optics, road signs, recognition, shape, color

I. INTRODUCTION

Road signs are very important part of transportation infrastructure. Their role is to regulate traffic, to navigate drivers to the certain parts of city or country and to insure safety of drivers. Transportation agencies have to control condition of all road signs regularly as part of service on each part of road. Then they have to provide necessary repairs, or replacement of damaged road signs. The function of proposed system is to solve this problem by automatic inventory of road signs. Every sign is characterized by its own shape and by representation of its color segments. Dimensions of signs are also specified but they can be different from the others, depending on type of road where they are installed. The second important property of road sign is its position. It can be relative distance from base point or it can be absolute position defined by GPS coordinates. It is possible to recognize existence of appropriate road sign on scanned scene. We can do it by comparing its geometric and colorimetric properties with properties of ideal road signs. It is also possible to determine conditions of this road sign effectively by parameters acquired from video log. Using the GPS receiver and laser distance meter we can determine sign's accurate position.

Important part of proposed system is Cambridge optical correlator. Cambridge optical correlator is able to evaluate similarity of two images. It uses Fourier optical transformation and autocorrelation of pictures composed by ideal road signs and scanned scene.

For optical correlator is necessary to preprocess pictures, with which we will work. Result of this preprocessing are binary images with geometric shapes. This modification is made by appropriate color filters and by segmentation process. Then the correlator is able to compare and signalize rate of similarity, with high rate of correct response. We can

determine position of road on logged scene based on Optical Correlator output. Then we can use it in another processing.

The condition of recognized sign is possible to compute using of sign geometric properties from video log. We know that video log is composed by sequence of several pictures. Based on sign movement among these pictures, it is possible to increase probability of correct recognition of sign.

In the first section of this paper we will talk about function of system core. In our case, the system core is the optical correlator. In second section we will describe possible way for video preprocessing for input of correlator. At last, there is presented the computing procedure for position and tilt of road sign. There are presented advantages and disadvantages of this system in conclusions. There is also presented a direction of research in the future.

II. OPTICAL CORRELATOR

Optical correlator is a device, which is able to compare two input pictures and determine their similarity. Optical correlator uses diffraction phenomena of optical lens. They provide Fourier transformation by optical way.

The main function parts of correlator are as follow:

Set of optical lens

Optical lens represent the most important part of optical correlator. They provide own Fourier optical transformation. At first, a picture is generated by space light modulator. The beams, which create this picture, go through the lens and create the picture of Fourier optical transformation on CMOS sensor.

Space light modulator (SLM)

Space light modulator is a device, which modulates light beam, on base of current information in input signal. The light beam directs on surface of space light modulator. Then it reflects current configuration of liquid crystals in SLM. The liquid crystals in every pixel of SLM are controlled by video signal through DVI or VGA interface. Control board addresses every pixel and sets tilt of liquid crystals. This controlling depends on video signal. Modulator is placed in front focal length of lens.

CMOS sensor

CMOS sensor is used for scanning optical correlation result. Final block of light beams is directed on surface of

CMOS sensor after processing the light beams by set of lenses. We can obtain final picture from CMOS sensor via Ethernet interface. First condition for correct function is connection between correlator and PC. The second condition is correct TCP/IP parameters settings. CMOS sensor is placed in rear focal length of optical lens.

Source of laser emission

The laser diode is used as source of laser emission. This laser diode emits coherent light. Light beam is led by optical fiber to Fourier optic set. Laser emission is used as carrying information source in optical correlator. It is modulated by configuration of space light modulator. Modulated optical signal is processed by Fourier optical transformation in form of adapted beams block.

The following picture shows the optical correlator output based on autocorrelation of input scene. Input scene is composed by reference sample and scanned sample. After autocorrelation the most dominant DC component, eventually its lateral peaks will be displayed. Lateral peaks represent accordance, or similarity of two parts of input scene. Using lateral peaks it is possible to determine where accordance or similarity takes place in scanned scene toward reference sample.

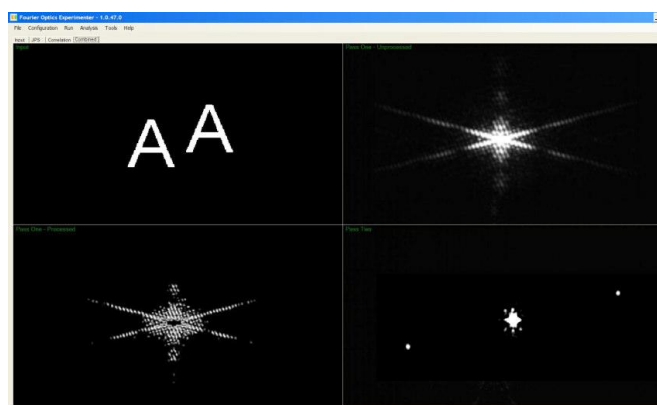


Fig. 1. Processing of input scene by optical correlator [7]

III. IMAGE PREPROCESSING FOR INPUT OF OPTICAL CORRELATION

Because the correlator works with help of coherent laser light, it is able to work with binary types of pictures. Therefore, we have to adjust the reference database and the scanned scene. This adjustment depends on characteristics of segments which have to be recognized. We can give example of road sign for speed limitation. In this case it is necessary to divide recognition and classification of road sign into two steps.

At first, it is necessary to evaluate existence of sign on logged scene. This evaluation can be made by comparing the scene with database of possible geometrical shapes. Those shapes represents contour of this road sign. In case of sign for speed limitation in Slovak Republic, it is necessary to recognize region with circle shape. After this operation we can make detailed classification by text inside the sign.

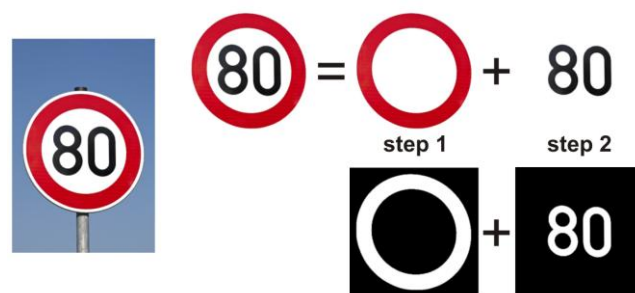


Fig. 2. Segmentation and thresholding of scene

For effective processing and for determining the rate of similarity, we have to remove unwanted parts from scene. In ideal case, the result of this operation should be contour of biggest object with determined colorimetric criteria. It can also be the whole scene with similar objects. At first step we have to convert color information of every pixel from RGB to HSV model. Colors in RGB model are defined by combination of three basic colors for each pixel. Then the final color is composed from parts of red, green and blue color. The HSV model defines colors by its hue, saturation and illumination. By using HSV model, we can manipulate with color in dimensions of particular hues. The necessary modification, which we have to make is to decrease saturation of color hues out of defined range for current recognized road sign. Result of this modification will be the picture with shining segments. Hue of these segments is in our defined range for this type of sign. Backgrounds saturation will be decreased.



Fig. 3. Decreasing of saturation for background colors [1]

Another step in preparation for input of correlator is the color thresholding. It is provided by using a software color filter. Software color filter provides operation, which removes all segments out of defined color range for current road sign. Colority of road sign in video log depends on environment conditions and on properties of the camera. The final color representation of the same road sign is different during the sunny weather and during the rain. For increasing the probability of correct thresholding we can create normalized model of processed colors. It is possible to reduce selected colors of three-dimensional RGB model into two-dimensional space by normalization of two color by third. [2]

IV. TOOLS FOR IMAGE PROCESSING

We can use the open source library named Open CV for image preprocessing. Open CV is a library of programming functions, mainly aimed at real time computer vision. Open CV was originally written in C but now has a full C++ interface and all new development is in C++. There is also a full Python interface in the library. Open CV is able to develop many kinds of graphic applications e.g. Human –

Computer Interaction, Segmentation and Recognition, Motion Tracking, Stereo and Multi – Camera Calibration etc.

Open CV has a modular structure which means that the package includes several shared or static libraries. [4]

ImgProc library

ImgProc is an image processing module of Open CV that includes linear and non-linear image filtering, geometrical image transformation, color space conversion, histograms, etc. This library contains the useful functions for our project. For example CvtColor() for image color space conversion and Threshold() for applying fixed-level threshold of selected color channel. [4]

V. ROAD SIGN ATTRIBUTES COMPUTATION

We can compute sign parameters based on the geometric configuration of the camera and the sign. We can use the rule of similar triangles

$$\frac{H}{W} = \frac{D}{S}, \quad (1)$$

where H is the height of the sign, D is the sign to camera distance. W is the sign height on the image and S is distance (in pixels) between the camera center point and sign bottom line. S can be computed from image coordinates, focal length and camera base point as

$$S = \sqrt{f^2 + (u_1 - u_0)^2 + (v_1 - v_0)^2} \quad (2)$$

Then the sign height is computed by following expression

$$H = \frac{W \cdot D}{f^2 + (u_1 - u_0)^2 + (v_1 - v_0)^2} \quad (3)$$

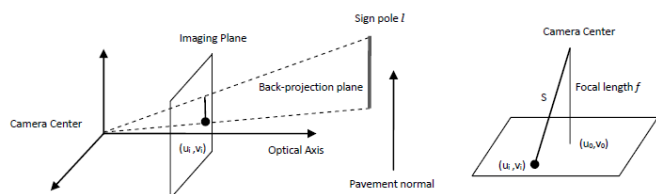


Fig. 4. Geometric configuration of road sign and camera for sign height computation [2]

By this way, we can also determine the second dimension of road sign. We can compute horizontal, and vertical tilt by comparing the proportion of real computed dimensions and proportion of ideal height and width of sign.

VI. USER INTERFACE AND DATABASE COMMUNICATION

The user interface for road signs inventory should be created in cooperation with Qt framework.

Qt framework allows us to write advanced applications and User interfaces once and deploy them across desktop and embedded operating systems without rewriting the source code saving time and development cost.

It contains many libraries for creating users interfaces, manipulating with many types of documents, for communication via the most common protocols, etc.

We also need storage for working data as sign shapes, sign pictograms, list of installed road signs and their conditions after inventory revision. Very useful database system for this purpose is MySQL created in Oracle corporation. MySQL is free for use and for distribution with each kind of

applications. It is platform independent and can be used on both personal computer and dedicated server. The big advantage is the direct support in web based programming languages which is useful for efficient sharing between two or more transportation agencies. [5]

VII. RECOGNITION AND INVENTORY PROCEDURE

We can describe whole process of road sign recognition and inventory using previous described parts:

1. Road scene is logged by digital camera.
2. Each image from video log is segmented into binary picture for every representing sign colors. For example, sign which describes maximum allowed speed has red circle shape with black number in its center.
3. All possible ideal shapes are prepared for comparison with images obtained from step 2.
4. Ideal shapes and binary input images are merged to one binary image separately. It depends on type of shape for input of optical correlator.
5. Correlator output for shape comparison is evaluated and position of sign candidate on image is declared.
6. Correlator output for number detection is evaluated and position of sign is confirmed.
7. Condition of road sign is obtained using dimensions from video log.
8. If sign is recognized correctly and its parameters are correct then it is logged into the database. In another case, the error message about wrong sign parameters is showed on screen.

VIII. CONCLUSION

Optical processors and correlators are more and more available for commercial sphere because of technology improvement. They have big power for use in transportation and its related sectors because optical lenses are able to provide the Fourier transformation in real time. By their implementation, there is saved high computing capacity of electronics systems. In our another research we will create detailed proposal and realization of described system. This system will be able to provide full automatic management of road sign inventory.

ACKNOWLEDGMENT

We support research activities in Slovakia / Project is co-financed from EU funds. This paper was developed within the Project "Center for excellence for integrated research and for using progressive materials and technologies in automotive electronics", ITMS 26220120055.



REFERENCES

- [1] M. Petruš, *Optical correlator at video driver assistance system*. Technical University of Košice, Košice, Slovakia, 2011.
- [2] Zhaozheng Hu, *Intelligent Road Sign Inventory (IRSI) with Image Recognition and Attribute Computation from Video Log*. Kyoto, Japan, 2011.
- [3] Guillem Vallicrosa Massaguer, *Road Sign Detection and Shape Reconstruction using Gielis Curves*. Bourgogne: Master in Computer Vision, 2011, ch. 4.
- [4] *Open CV library*, published on: <http://opencv.willowgarage.com/wiki/>.
- [5] *Qt framework*, published on: <http://qt.nokia.com>
- [6] Š. Toth, *Rozpoznávanie dopravných značiek a ich použitie v mapových aplikáciách*. GIS, Ostrava, 2011
- [7] *Cambridge optical correlator*, published on <http://www.cambridgecorrelators.com/technology.html>

Sensorless vector control of induction motor with neural speed estimator

¹Peter NGUYEN (3rd year), ²Marek VACEK (1st year)
Supervisor: ³Jaroslava ŽILKOVÁ

^{1,2,3}Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

¹peter.nguyen@tuke.sk, ²marek.vacek@tuke.sk, ³jaroslava.zilkova@tuke.sk

Abstract—Development of the speed estimator based on an artificial neural network for use in the sensorless vector control of induction motor are described in the paper. The structure of neural speed estimator is simple and requires only present and past values of two variables. Neural network was trained off-line based on data obtained from the simulation of vector control. This type of speed estimator is verified by simulation and some tests to changes of angular speed and load torque are done.

Keywords—angular speed, estimator, induction motor, neural network, sensorless vector control.

I. INTRODUCTION

Induction motor drives are widely used in different industrial branches because of its simple construction. Induction motor can be controlled either by a scalar control, direct torque control, or vector control. If high accuracy and dynamic control is required, then is necessary to use vector control of induction motor.

Nowadays the sensorless control of electrical drives is extensively investigated, because elimination of speed sensor provides several advantages, such as cost and size reduction of electrical drives, higher reliability and less maintenance requirements.

Various methods have been proposed for the speed estimation of induction motor, such as open-loop estimator [1], estimator using saturation of third harmonic voltage, model reference adaptive systems (MRAS) with different kind of observers [2], [3], [4], adaptive observers like Luenberger observer, Kalman filter or Sliding mode observer [5], signal injection techniques [6] and estimators using artificial intelligence [7], [8].

Several neural speed estimators have been proposed in [10], [11] and [12]. In [10] authors designed two neural estimators. First estimator was proposed in stationary reference frame (α, β) and has three inputs (suitable modified signals of stator current and voltage components). However for a good accuracy of speed estimation is necessary to have a small sampling time. The second estimator was proposed in synchronously rotating reference frame and has four inputs, present and past values of stator voltage and current components. In this case for a good estimate of the speed is sufficient 1 ms sampling time. In [11] is designed similar

neural speed estimator, but has two additional inputs, i.e. torque producing current component. Authors in [12] proposed speed estimator, which has eight inputs, i.e. stator current and voltage components expressed in stationary reference frame. This neural speed estimator is more complex than above mentioned and for a good estimate requires a small sampling period.

The present paper discusses the design of neural estimator similar to that mentioned in [10], but in a different structure of vector control of induction motor.

II. NEURAL SPEED ESTIMATOR

A. Mathematical model of induction motor

The mathematical model of an induction motor (IM) in a synchronously rotating frame ((x,y), labeled by subscript k) can be described by equation as follows.

The stator and rotor voltage equations:

$$\mathbf{u}_{1k} = R_1 \mathbf{i}_{1k} + \frac{d\mathbf{\Psi}_{1k}}{dt} + j\omega_k \mathbf{\Psi}_{1k} \quad (1)$$

$$0 = R_2 \mathbf{i}_{2k} + \frac{d\mathbf{\Psi}_{2k}}{dt} + j(\omega_k - \omega) \mathbf{\Psi}_{2k} \quad (2)$$

The stator and rotor flux linkages in rotating reference frame (x,y) are:

$$\mathbf{\Psi}_{1k} = L_1 \mathbf{i}_{1k} + L_h \mathbf{i}_{2k} \quad (3)$$

$$\mathbf{\Psi}_{2k} = L_2 \mathbf{i}_{2k} + L_h \mathbf{i}_{1k} \quad (4)$$

Where R_1 and R_2 are stator and rotor resistance, i_1 a i_2 are stator and rotor current, L_1 , L_2 and L_h are stator, rotor and mutual inductance, ω_k is angular speed of synchronously rotating reference frame (x,y) and ω is rotor angular speed.

The electromagnetic torque and rotor speed are related by the dynamic equation:

$$T_e - T_L = \frac{J}{p} \frac{d\omega}{dt} \quad (5)$$

$$T_e = \frac{3p}{2} \text{Im}[\mathbf{\Psi}_{1k}^c \mathbf{i}_{1k}] \quad (6)$$

Where J is the moment of inertia, p is number of pole pairs, T_e and T_L are electromagnetic and load torque respectively, superscript c denotes the conjugate value of variable.

B. Design of neural speed estimator

Neural networks represent very good nonlinear function approximators. If there are suitable data describing the behavior of system, properly designed and trained neural network can be used as an estimator.

Neural speed estimator design for sensorless vector control of IM has several parts. First it is necessary to obtain data that capture the most possible operating conditions of controlled drive. Next, it is necessary to select appropriate inputs for neural network, from which will determine the desired output variable, in our case angular speed. And it is also necessary to choose an appropriate network structure, i.e. how many inputs

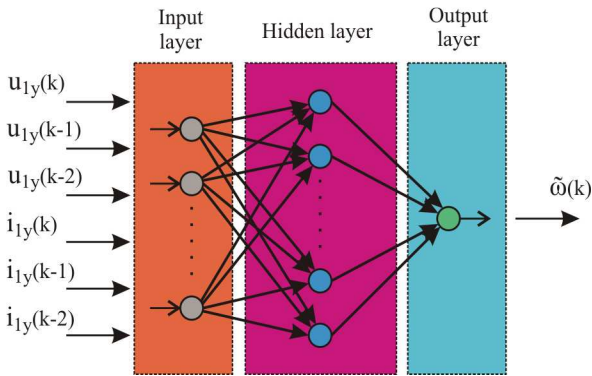


Fig.1. Multilayer feedforward neural network for angular speed estimation

will have the neural network, the number of hidden layers and hidden nodes, types of activation function in each layer, etc. If this is all done, we begin to train the neural network by certain learning algorithm. Then trained artificial neural network can be tested in vector control of induction motor.

A multilayer feedforward neural network with one hidden layer is used to estimate angular speed. Network inputs are the present and past values of stator voltage u_{1y} and stator current i_{1y} component. The output layer consists of one neuron representing the angular speed ω . The number of hidden neurons in the hidden layer is chosen by a trial-and-error technique, hence in the hidden layer are nine hidden neurons. In hidden layer is used hyperbolic tan (tansigmoid) activation function and in output layer linear (purelin) activation function. And therefore is used neural network with 6-9-1 structure. Neural network was trained using Levenberg-Marquardt algorithm in an off-line mode (supervised learning) and for learning have been used samples obtained from the simulation of vector control of induction motor.

Angular speed estimation can be realized using the feedforward neural network approximation of the relationship:

$$\hat{\omega}(k+1) = \mathbf{f}[u_{1y}(k), u_{1y}(k-1), u_{1y}(k-2), i_{1y}(k), i_{1y}(k-1), i_{1y}(k-2), \mathbf{w}] \quad (7)$$

Where \mathbf{f} is transfer function and \mathbf{w} is vector of weight and threshold coefficients.

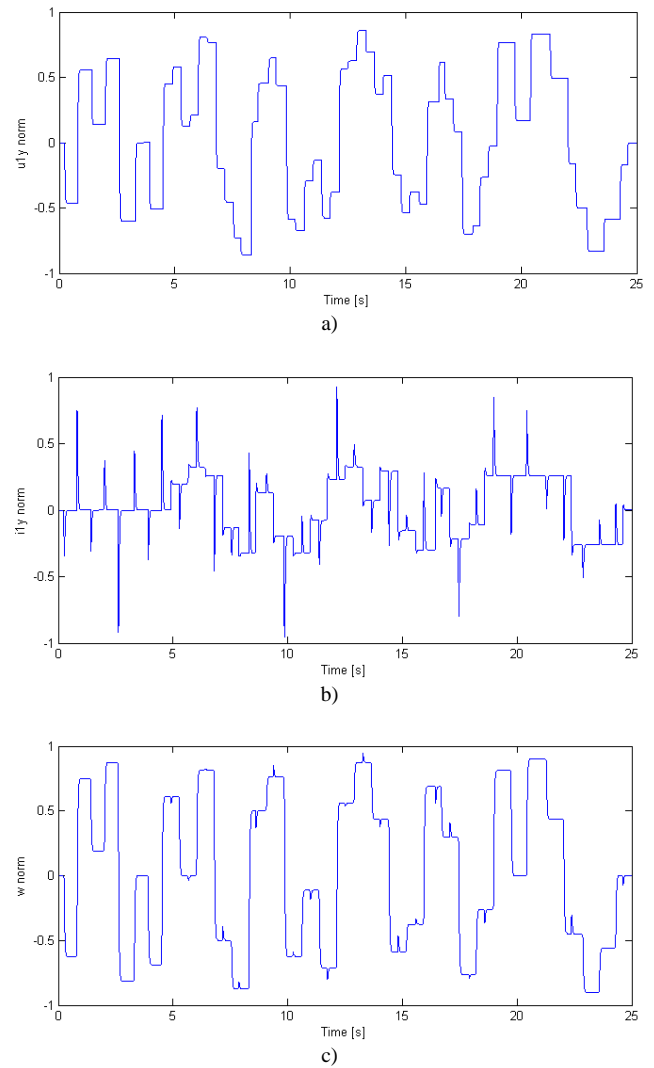


Fig. 2 Set of input/target data, all normalized: a) u_{1y} , b) i_{1y} c) ω

It should be noted that all variables used to train neural network are normalized to the range (-1,1) as follows:

$$\text{norm } u_{1y} = 1/200; \quad \text{norm } i_{1y} = 6,5; \quad \text{norm } w = 1/160.$$

III. SIMULATION RESULTS

The neural speed estimator properties were verified in the structure of vector control. State control of induction motor was proposed using pole placement method, which ensures a quick response to the desired value of speed without overshoot and also fast fault compensation [13]. The induction machine parameters are given in Table I.

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. 3.

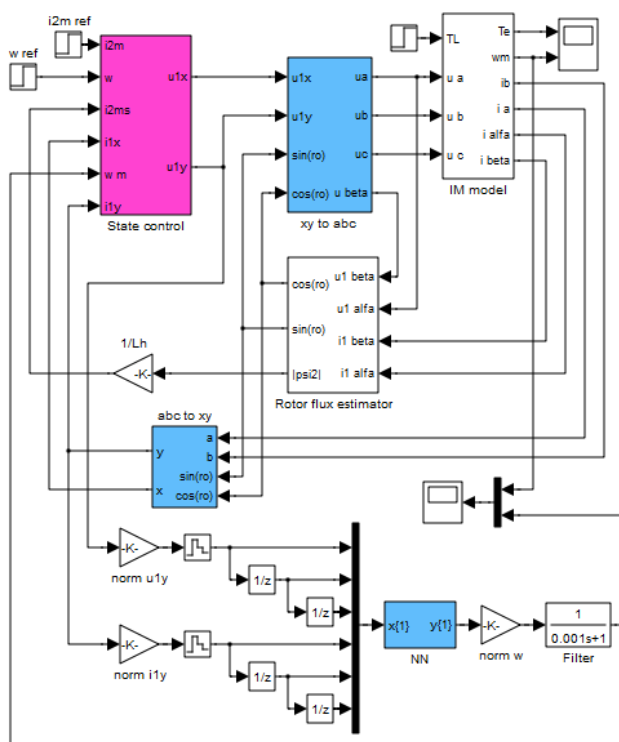


Fig. 3 Block diagram for testing the neural speed estimator in Simulink

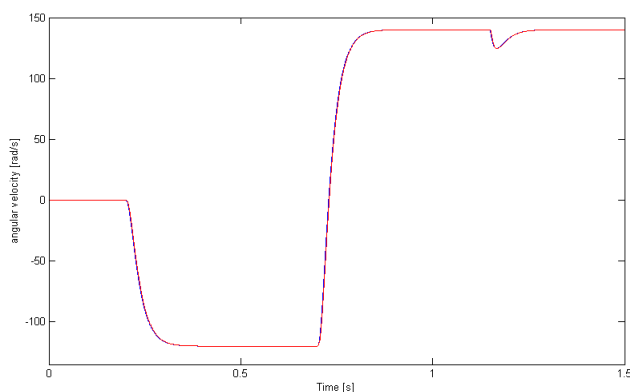


Fig. 4 Speed response during reversion from -120 to 140 rad/s

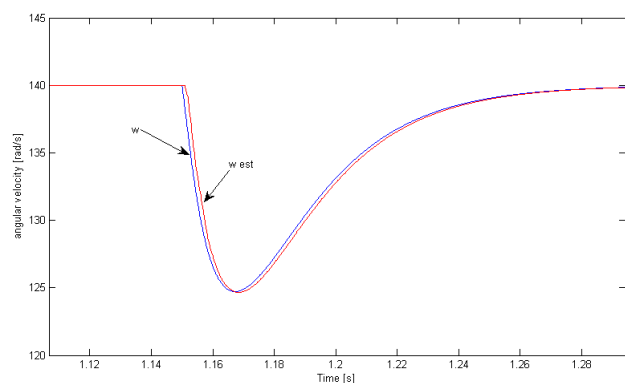


Fig. 5 Deatail waveforms of actual and estimated angular speed at the time of step change of load torque

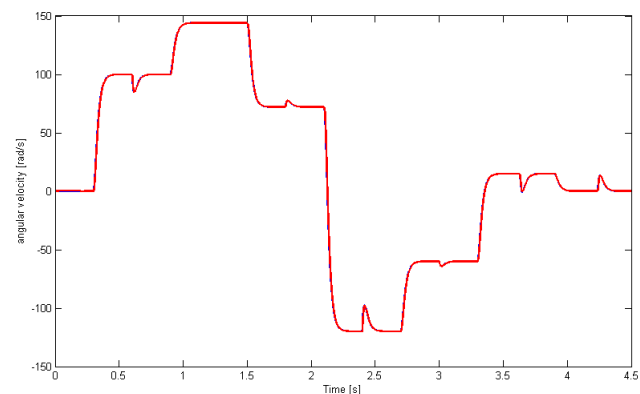


Fig. 6 Simulation results of the IM drive with neural speed estimator for random step changes speed and torque

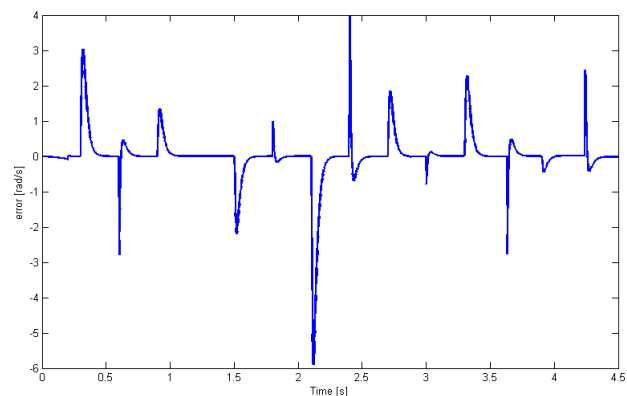


Fig. 7 Error between actual and estimated angular speed for random step changes speed and torque

Fig. 4 shows speed response of the vector-controlled induction motor and speed estimator based on artificial neural network during speed reversion from -120 rad/s to 140 rad/s. Fig. 5 shows detail of real and estimated speed during step change of load torque at the time 1.15 s from 0 to nominal load torque 5Nm. From the figure is seen that the estimated angular speed is delayed compared to the actual. This is due to the fact that on the output of the neural estimator is used filter.

Fig. 6 shows speed response of the vector-controlled induction motor and neural speed estimator for the random step changes of the speed and torque. From the figure is seen that the neural speed estimator achieves good performance. This is confirmed by figure 7 which shows the deviation between the actual and estimated angular speed. The largest deviation is about 6 rad/s^{-1} .

IV. CONCLUSION

This article has demonstrated excellent approximation properties of neural networks. For vector controlled induction motor was designed neural speed estimator, which consists of a feedforward neural network with one hidden layer. Network inputs are present and past values of stator voltage u_{1y} and stator current i_{1y} component. Simulation results confirm good accuracy of estimation of angular speed, both at low and high speeds, and also at various loads. It follows that a well designed neural speed estimator can be used for sensorless control of induction motor.

ACKNOWLEDGMENT

The support provided by the grant VEGA 1/0006/10 is kindly acknowledgment.

REFERENCES

- [1] J. Holtz, "Sensorless control of induction machines – with or without signal injection?," *IEEE Transaction on industrial electronics*, vol. 53, no. 1, pp. 7-30, 2006.
- [2] S. Meziane, R. Toufouti, H. Benalla, "MRAS based speed control of sensorless induction motor drives," *ICGST International Journal on Automatic Control and Systems Engineering, ACSE*, vol. 7, Issue 1, pp. 43-50, May 2007.
- [3] T. Orłowska-Kowalska, M. Dybkowski, "Stator-Current-Based MRAS Estimator for a Wide Range Speed-Sensorless Induction-Motor Drive," *Industrial Electronics, IEEE Transactions on*, vol. 57, no. 4, pp. 1296-1308, April 2010.
- [4] M. Žalman, M. Abelovský, "Pozorovatele stavových veličin bezsnímačových servopohonov s AM," In: *AT&P Journal Plus*, no. 6: Mechatronika, pp. 21-32, 2005.
- [5] Y. Zhang; Z. Zhao; T. Lu; L. Yuan; W. Xu; J. Zhu; , "A comparative study of Luenberger observer, sliding mode observer and extended Kalman filter for sensorless vector control of induction motor drives," *Energy Conversion Congress and Exposition, 2009. ECCE 2009. IEEE* vol., no., pp. 2466-2473, 20-24 Sept. 2009.
- [6] G. Qiang, G. M. Asher, M. Sumner, P. Makys, "Position Estimation of AC Machines Over a Wide Frequency Range Based on Space Vector PWM Excitation," *Industry Applications, IEEE Transactions on*, vol. 43, no. 4, pp. 1001-1011, July-aug. 2007.
- [7] S. M. Gadoue, D. Giaouris, J. W. Finch, "Sensorless control of induction motor drives at very low and zero speeds using neural network flux observers," *IEEE Transaction on industrial electronics*, vol. 56, pp. 3029-3039, 2009.
- [8] L. Yi, H. Mei, "Speed sensorless IM control system based on MRAS and NN flux observers," *3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, pp. V5-163-V5-167, 2010.
- [9] P. Vas, *Artificial-intelligence-based electrical machines and drives*. New York: Oxford University Press, 1998, pp. 349–369, ISBN: 0 19 859397 X.
- [10] J. Jovanković, M. Žalman, "Application of the virtual sensors based on the artificial neural networks sensorless drive," In: *AT&P Journal*, vol. 11, no. 2, pp. 67-70, 2004.
- [11] P. Gírovský, J. Timko, J. Žilková, V. Fedák, "Neural estimators for shaft sensorless FOC control of induction motor," *EPE-PEMC, International conference*, Macedonia, 2010, pp. T7-1-T7-5.
- [12] M. Kuchar, P. Brandstetter, M. Kaduch, "Sensorless induction motor drive with neural network," *35th Annual IEEE Power Electronics Specialists Conference*, Germany, pp. 3301-3305, 2004.
- [13] L. Zboray, F. Ďurovský, J. Tomko, *Regulované pohony*. Košice, Viena, 2000, pp. 205–225. ISBN: 80-88922-13-5.

Sources, use and measurement of electromagnetic field

¹Ján Zbojovský (1st year), ²Marek Pavlík (1st year)
Supervisor: ³Alexander Mészáros

^{1,2,3}Dept. of Electric Power Engineering, FEI TU of Ko-ice, Slovak Republic

¹jan.zbojovsky@tuke.sk, ²marek.pavlik@tuke.sk, ³alexander.meszaros@tuke.sk

Abstract This article shows the basic properties of electromagnetic fields as sources of electromagnetic radiation, and methods of measurement of electromagnetic fields.

Keywords electromagnetic fields, properties of electromagnetic fields, sources of electromagnetic fields

I. INTRODUCTION

People are continuously exposed to electromagnetic fields. This environment consists of natural electromagnetic radiation and man-made electromagnetic fields that are produced either intentionally or as a side effect of the operation of equipment. Each country has made its own standards in terms of exposure to electromagnetic fields, and these standards are based on guidelines set by an international commission for protection against non-ionizing radiation (ICNIRP). Effects of exposure to electromagnetic fields have been a subject of many researches.

II. NATURAL AND MAN-MADE SOURCES OF ELECTROMAGNETIC RADIATION

The electromagnetic field is a physical field in which electric and magnetic forces operate in space.

An natural electromagnetic environment comes from terrestrial and extraterrestrial sources such as radiation from space and the sun, or from electrical discharges in the Earth's atmosphere. Compared to man-made fields, natural fields are extremely small at radio-frequencies [4].

In the case of low frequencies (below 30MHz) electromagnetic radiation is caused by lightning during a storm. In many cases, it is lightning between clouds, but lightning between cloud and ground is also dangerous. The intensive current impulse associated with the discharge generates wideband electromagnetic impulses, which are spread over large distances in the conductive ionosphere waveguide and on the Earth's surface.

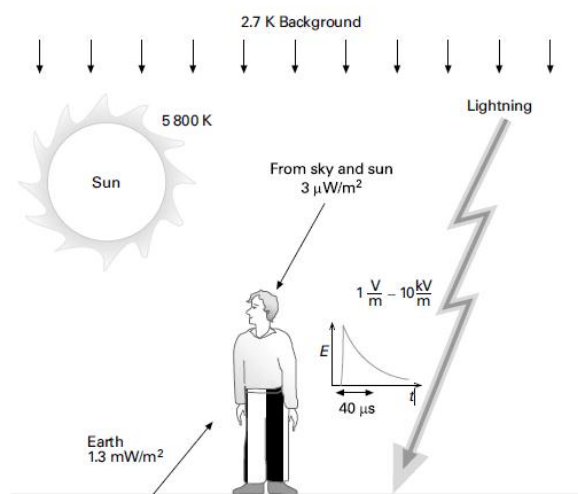


Fig.1 Terrestrial and extraterrestrial sources of high-frequency radiation [1]

For high frequencies (above 30MHz) natural electromagnetic fields originate from the radiation of the Earth and extraterrestrial sources, mainly from the sun and the sky itself. It should be noted that only for the frequency range from 30MHz to 30GHz, electromagnetic waves penetrate the atmosphere efficiently. In case of frequencies below 30 MHz, waves are reflected back into space, and for frequency higher than 30GHz attenuation occurs.

Man-made electromagnetic fields spread from sources such as power lines, telecommunication, radio-television and many other devices such as electric motors, railways, and other welding equipment [1].

III. USE OF MAN-MADE SOURCES OF ELECTROMAGNETIC RADIATION

As mentioned above, man-made electromagnetic fields are created either intentionally or as a product of the operation of electrical equipment. Next it will be shown an overview of the use of fields [1].

Telecommunication and broadcasting the main objective is to spread the transmission of electromagnetic energy into the space around the transmitter. Generally, antennas of transmitting stations are the strongest permanent sources of radio frequency energy radiated into free space intentionally. Regarding the impact of radiation on humans, the critical

groups are those, who are working in broadcast towers.

*Mobile and wireless communication devices*ó wireless communication devices are used today in nearly all parts of modern society. A brief overview of the development of radio systems:

1G systems were based on analog transmission and worked at a frequency of 450MHz and 800/900MHz using frequency modulation.

2G systems based on digital mobile communications system. Currently, there are many different systems. In Europe, America and some parts of Asia, GSM is the dominant system worked at 900 and 1800MHz. Japan has developed its own system worked at 1,5 GHz. 2,5 G systems were associated with increasing transmission speed due to the development of Internet use.

3G systems represent the latest digital mobile communication technology. Systems work at frequencies of 1900 and 2200MHz.

The successor to 3G systems is 4G, working at 2 and 6 GHz and reaching the speed of 100Mbit/s in the external environment and 1Gbit/s in the indoor environments.

*Microwave communication links*ó two highlyó directional antennas are used to provide communication lines in various applications such as mobile communications in public and private communication equipment. The system is used for transmission over long distances. An example is satellite transmission.

*Use in medical applications*ó electromagnetic fields were primarily used in diathermy. Commonly used are two types: short waves and microwaves. Intensity of exposure is sufficient to cause the temperature rise of tissues. Another possibility is the use for electromagnetic resonance. It is used to create images of apparatus and structures inside the human body.

*Use in industrial and domestic applications*ó electric and magnetic fields are used in the processing of various materials such as those used in heating and sometimes forming a plasma discharge in the material. It is impossible to oversee safety problems because RF power is high. Therefore a good shielding is necessary to prevent radiation leakage.

*Dielectric heating*ó high-frequency dielectric heating is one of the largest sources of radiation. It is used in the welding of plastics. It is used to 20kV voltage and frequencies of 5 to 40 MHz.

*Induction heating*ó heating of electricallyóconductive material is performed by eddyócurrents that are induced by an electromagnetic field. It is used in soldering and welding. The frequency varies depending on the size of the heated material. Induction heating is also used in household induction cookers.

TABLE I
OPERATING FREQUENCY OF INDUCTION HEATING [1]

Frequency (kHz)	The size of the heated material
5-30	Strong materials
100-400	Small parts
480	Microscopic pieces

An intense electric field is used to produce a plasma discharge, which is used in the manufacture of semiconductors. The operating frequency is 13,56 MHz and power in the range of 100-watts to kilowatts. Measurements have shown that exposure of workers at a distance of more than 10 cm is relatively small.

*Microwave heating and drying*ó microwave energy is used for heating and drying of various materials such as food, construction materials, health and so on. The used frequency is 2450 MHz, and 915MHz in some countries.

*Security applications (radar, navigation) and new technologies*ó radar systems usually operate at frequencies from 500MHz to 15GHz, and also exists systems, which are operating at frequencies above 100GHz. Radar employed in air transport is used to detect air conditions, condition of the runways and etc. The antenna rotates continuously resulting in a radiated power in one direction which is relatively small. The operating frequency of the radar is 2,8 GHz. UWB technology is mainly used in sensing devices such as trafficóradar. The UWB technology allows a frequency range between 1,99 and 10,6 GHz and between 22 and 29GHz.

*Wireless LANs*ó together with the development of mobile communications continuous development of wireless local area networks with a relatively short range between the user and the access point. They are called ad-hoc systems, set up in households, hotels, cafes, schools, as hot spots, and usually used to connect to the Internet.

*Wireless transmission of electricity*ó The concept is based on the transmission of electricity produced in solar panels in space and its subsequent transfer to Earth. Satellite receivers on earth would give 1 and 6 GW of power and operate at a frequency of 2,45 or 5,8 GHz. The higher rate, of 5,8GHz was proposed for the transfer. This system is under development and testing.

*The exposure systems used in laboratory conditions*ó the main task of these systems is to ensure a well-defined electromagnetic exposure to the studied subject. This includes all exposure parameters and their changes in space and time. In addition, they must satisfy certain criteria in order to prevent other influences associated with the system itself and studied the subject.

These systems are divided into two groups [3]:

In vitro exposure systems: an important factor in the design of exposure systems "in vitro" is a coupling between electromagnetic fields and media. For "in vitro" systems were developed several technical solutions.

In vivo exposure systems: should satisfy the same criteria as in vitro but with regard to the needs of animals.

Both of these systems examine the impact of exposure to a biological organism.

IV. MEASURING OF ELECTROMAGNETIC FIELD

RF sources allow formation of electric and magnetic fields, which can bind directly to humans, induced fields and currents in the human body. There are many factors that affect the measurement and its use in exposure assessment, because there is a range of approaches to exposure assessment. The components of the electromagnetic field can vary in space and time with respect to their size and direction. The objective of the measurement is to obtain information about these values and demonstrate methods to assess compliance with guidelines, standards and regulations, and also for the exposure assessment for health-related research [1].

Principles of measurement

Measuring devices with sufficient accuracy must be able to measure necessary values with respect to the characteristic properties of the signal. It must also have sufficient sensitivity and frequency range. Measurement results can be affected by environmental parameters such as temperature, humidity, the device itself or interference. Also it may be the startúp of resistance equipment, interconnecting cables, other effects of magnetic fields as a component in measuring the electric field and vice versa. When selecting a measuring device is necessary to take into account a number of key factors such as instrument response, peak power, sensor limits, dynamic range and the ability to measure near and far fields, depending on the measurement conditions in the field [2].

The characteristics of electromagnetic fields

During the measurements the following parameters should be taken into account:

- ó Modulation signal ó depending on the time and frequency
- ó Wave propagation ó in place of measurement are various reflected waves.
- ó Signal attenuation as a statistical variation in amplitude due to the multidirectional spread between the source and the point of measurement.
- ó The spectrum of radiation generated by source, which is the spatial distribution of an electromagnetic field. In the near field, the field distribution varies with distance from the source. In the far field there are no significant changes dependent on the field.
- ó Frequency spectrum of sourceó energy can be distributed through several frequencies. For example, UWB sources have energy distributed in ranges from 3,1 to 10,6 GHz.
- ó Impedance of the field describes the quantity of energy associated with electric or magnetic field at each point of interest.
- ó Polarization of a field can be constant (linear polarization) or rotary (elliptical polarization).
- ó The way of the field distribution in space.
- ó Spatial distribution of the field depending on the location of the source.
- ó An environment located between the source and the point of measurement.

Measurement of the Faró Field

Far field represents an area away from the source of the electromagnetic field. Electric E and magnetic H component is essentially independent of distance from the source. The field has a mainly plane wave character, which means the uniform distribution of the electric and magnetic field in the plane vertical to the direction of propagation. The components E and H are vertical to each other [3].

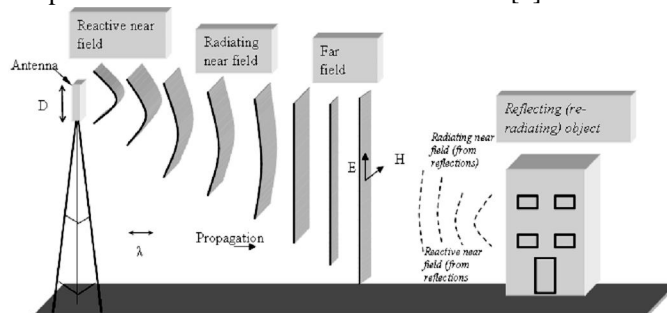


Fig.2 Near and Far Field [1]

*Measurements in case of one source*ó the measurement can be performed with simple equipment. It measures the part of a field, either electric or magnetic, according to needs.

Measurement in the case of multiple sources ó must be used special practises. The accuracy of measurement requires attention to the frequency, polarization, modulation and time of turning off or turning on each source. Also, interference can affect accuracy of measurement.

Measuring Nearó Field

Spatial distribution of electric and magnetic components are independent of one another. Amplitudes of E and H components vary depending on the distance of the source and therefore must be measured at each point of interest [3].

Instrumentation for measuring electromagnetic fields: Portable RF meters are relatively easy for measuring electric and magnetic fields and subsequent assessment of compliance with the exposure conditions. Most of the available measuring equipment is broadband.

The main components of broadband devices are:

- ó Sensoró antenna or detection device which produces an audio signal proportional to the size of the field strength.
- ó Data flowó metallic or optic cable used to transmit power from the antenna to the device for processing and displaying data
- ó Equipment for processing and displaying data used for the processing of measured data, filtering, and digitize summary to display data to display [1].

The antenna for measuring the electric field consists of one or two dipoles. Isotropic probes contain three mutually orthogonal dipoles and derive a vector summation of their outputs to give a response independent of probe orientation. The antennas in magnetic field probes are usually three mutually orthogonal loop or coil elements.

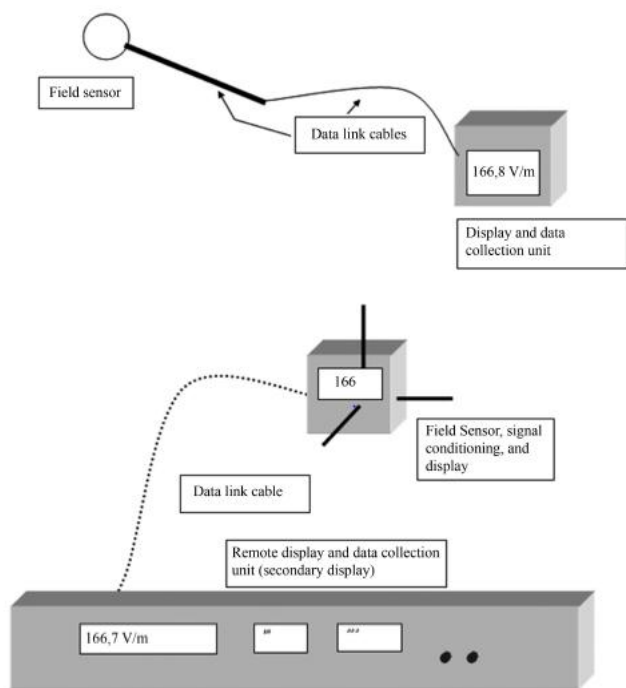


Fig.3 Schematic of broadband instrument [1]

Spectrum analyzer

The device is designed to measure the nearly fields. The dimensions of the measuring probe should be a fragment of the wavelength of the highest operating frequency. The response of the sensor should be isotropic, independent of the orientation and polarization of the measured fields. Sensor with nonó isotropic response is useful if the polarization of measured parameters E and H are known and the sensor can be rotated in the direction of polarization. These devices are tunable over a wide frequency range and can be used to display changes in a specific part of the amplitude spectrum.



Fig.4 Spectral Analyzer- measurement kit [1]

List of parameters to be considered when using a spectrum analyzer:

- ó Frequency responseó this is the bandwidth in which the analyzer operates.
- ó Resolution bandwidthó is a band in which the analyzer measures the field strength at a certain frequency. The measurement is performed by Gaussian filter. Insufficient resolution may degrade the readability of the field strength.
- ó Number of pointsó is the number of discrete frequencies at which measurements are performed.

Currently, there are many other devices for measuring exposure and to assess the impact on human health and etc.

Table 3 shows some representative exposure data in the positioning of RF induction heaters. The magnetic field strength varied from 0,2 to 20A.m⁻¹ and the electric field from 10 to 1600 V.m⁻¹ in the position of head and hands [1].

TABLE II
MEASUREMENT OF ELECTRIC AND MAGNETIC FIELDS AT THE POSITION OF THE OPERATOR OF AN INDUCTION HEATER [1]

Frequency (kHz)	Magnetic field strength (A.m ⁻¹)		Electric field strength (V.m ⁻¹)	
	Head	Hands	Head	Hands
484	1,44	-	650	500
743	0,88	0,72	160	32
394	1,52	12,88	168	70
300	0,24	0,24	16	8
630	1,28	0,80	35	23
785	14,64	9,92	929	36
790	7,04	8,64	413	16
434	20,48	20,48	1192	646
500	8,48	-	192	64

V. CONCLUSION

This article provides a basic overview of the issue of electromagnetic fields. In the future, it is recommended to analyze the impact on building materials and building themselves of the distribution of the electromagnetic field and the execution of related measurements, and also to analyze the effects of electromagnetic fields on biological organisms.

ACKNOWLEDGMENT

This publication is the result of the Project implementation: Protection of the population of the Slovak Republic from the effects of electromagnetic fields, ITMS: 26220220145, Activity 2.1 supported by the Research & Development Operational Programme funded by the ERDF .



We support research activities in Slovakia / Project is cofinanced from EU funds.

REFERENCES

- [1] Paolo Vecchia, Rüdiger Matthes, Gunde Ziegelberger, James Lin, Richard Saunders, Anthony Swerdlow, *Exposure to high frequency electromagnetic fields, biological, effects and health consequences*, INCIRP 2009, ISBN 978-3-934994-10-2.
- [2] Christopher J. Portier, Ph.D, Mary S. Wolfe, Ph.D, *Assessment of Health Effects from Exposure to Power-Line frequency Electric and Magnetic fields*, 1998, NIH Publication No. 98-3981.
- [3] *Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic field*, INCIRP, 1998 Health Physics Society
- [4] David M. Cook, *The theory of the electromagnetic field*, Prentice_Hall, New Jersey 1975 , ISBN 0-13-913293-7

Specificity of DC consumers in independent power supplies based on an asynchronous generator

¹Iurii ZACHEPA, ²Ján PERDULAK(2nd year), ³Tibor VINCE

¹Institute of Electromechanics, Energy Saving and Control Systems,
Kremenchuk Mykhailo Ostrogradskyi National University, Ukraine

^{2,3}Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

¹zip1981@ukr.net, ²jan.perdulak@tuke.sk, ³tibor.vince@tuke.sk

Abstract—The paper presents a mathematical model of an autonomous power system based on an asynchronous generator with a constant current customers that are connected through uncontrolled rectifier. For the purpose of the universality the rectifier and the consumers are represented by separate units, and a mathematical model is composed by the corresponding compound of the structural elements. The loading modes of asynchronous generator defined by its overload capacity and revealed the specifics of the physical phenomena in the system “asynchronous generator - capacitor bank - the rectifier” were investigated.

Keywords— asynchronous generator with capacitor excitation, DC motors, uncontrolled rectifier.

I. INTRODUCTION

In case of energy instability absence of centralized energy supply can be completed by autonomous power plants of low power, even in far located or economically undeveloped areas. Three main factors should be considered in systems of independent energy (ISE) – power of energy sources, the maximum power consumption and time of consumption the power [1]. Asynchronous machines in generator mode with capacitive excitation are widely used as a source of electricity. However, there are very large group of DC consumers. In this case asynchronous generator (AG) is connected to the frequency converter or rectifier (R). In small wind and hydro power plants, in heating systems of individual housing ISE energy converts into the heat and it is not important the type of electricity – AC or DC. It is known [2] that the using of rectifier at the output of the generator reduces the cost of the required capacitor excitation and increase the rigidity of the external characteristics of AG. But in the technical literature this issue is given little attention.

II. OBJECTIVE

Development of mathematical apparatus for asynchronous generator with capacitor excitation - rectifier - load (AG-BC-R-L) and study of the specific physical phenomena in such complex.

III. MATERIALS AND FINDINGS

A. Mathematical model of the rectifier (R)

Currently, there are a large number of modeling schemes

with converters [3, 4].

Method of switching functions was used in the model of the rectifier. This method of modeling the rectifier was chosen due to already pre-developed mathematical model AG-BC presented in [5].

The following mathematical model is referred to a six-pulse rectifier bridge circuit (fig. 1).

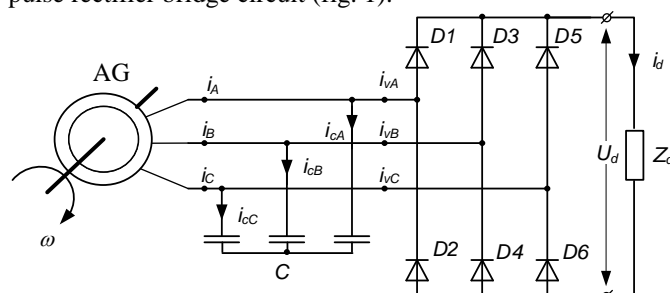


Fig. 1. Asynchronous generator with uncontrolled rectifier bridge

The path of the rectified current i_d through the diodes is calculated by using the switching functions F_i . The front functions F_1, \dots, F_6 are the same to beginning front intervals of the pairs of diodes $D_1 - D_6, \dots, D_5 - D_6$, and the trailing edge for each of these functions coincides with the leading edge of the follow. Numbers and diagram of concurrent diodes are shown in fig. 2. More details for this method of determining the switching functions is presented in [6].

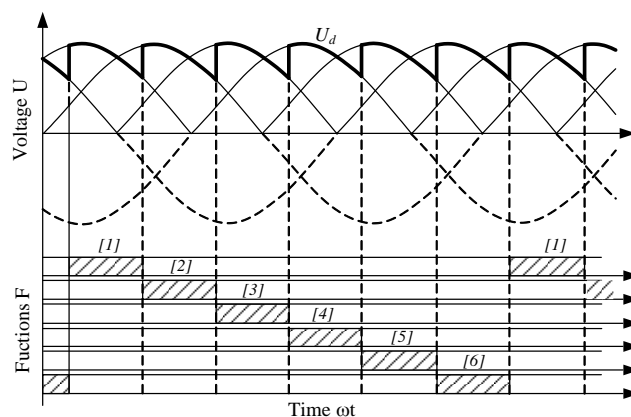


Fig. 2. Time diagrams of the converter

Amplitude and mean values of voltage for this rectifier scheme are determined as:

$$U_{dm} = \sqrt{6} \left[U_{Ac}(F_1 - F_4) + U_{Bc}(F_3 - F_6) + U_{Cc}(F_5 - F_2) \right] \quad (1)$$

$$U_d = \frac{3\sqrt{6}}{\pi} \left[U_{Ac}(F_1 - F_4) + U_{Bc}(F_3 - F_6) + U_{Cc}(F_5 - F_2) \right] \quad (2)$$

The average value of the rectified current:

$$U_d = \sqrt{\frac{3}{2}} \left[i_{Aa}(F_1 - F_4) + i_{Bb}(F_3 - F_6) + i_{Cc}(F_5 - F_2) \right] \quad (2)$$

B. Mathematical model of the actively-inductive loading block (RL)

According to the second law of Kirchhoff, the instantaneous values of phase voltage on loading looks like:

$$L_d \frac{di_d}{dt} = U_d - i_d R_d, \quad (3)$$

where L_d – inductance of loading, R_d – resistive impedance of loading.

C. Mathematical model of the motor loading (DCm)

Mathematical model of DC motor is represented as a system of differential equations:

$$\begin{cases} U_d = L_\Sigma \frac{di_d}{dt} + \omega k F + i_d R_\Sigma \\ J \frac{d\omega}{dt} = M_m - M_c \end{cases}, \quad (4)$$

where L_Σ, R_Σ – total inductance and resistance of anchor chain; k, F – constructive constant of the motor and magnetic flux; J – moment of inertia DCm; $M_m = i_d k F$ – moment on the motor shaft.

Imitation of different types of consumers is effected by creation of corresponding moment of resistance on the asynchronous motor shaft:

$$M_c = M_0 + (M_m - M_0) \left(\frac{\omega}{\omega_m} \right)^\alpha, \quad (5)$$

where M_m – rated torque of the system; α – coefficient depends on the consumer type, $M_0 \approx 0,05 - 0,1 M_m$ – an idling torque of DC motor.

As a result a general diagram of the system AG-consumer is brought on fig. 3, where M_{out} is a outsourcing torque of mechanical energy (combustion engine, hydro- or windmill); BC are condensers of excitation; AG is an asynchronous generator; S1 is the guided key for the type selection of consumer; BTM_c is a block of task of moment on the shaft of DCm.

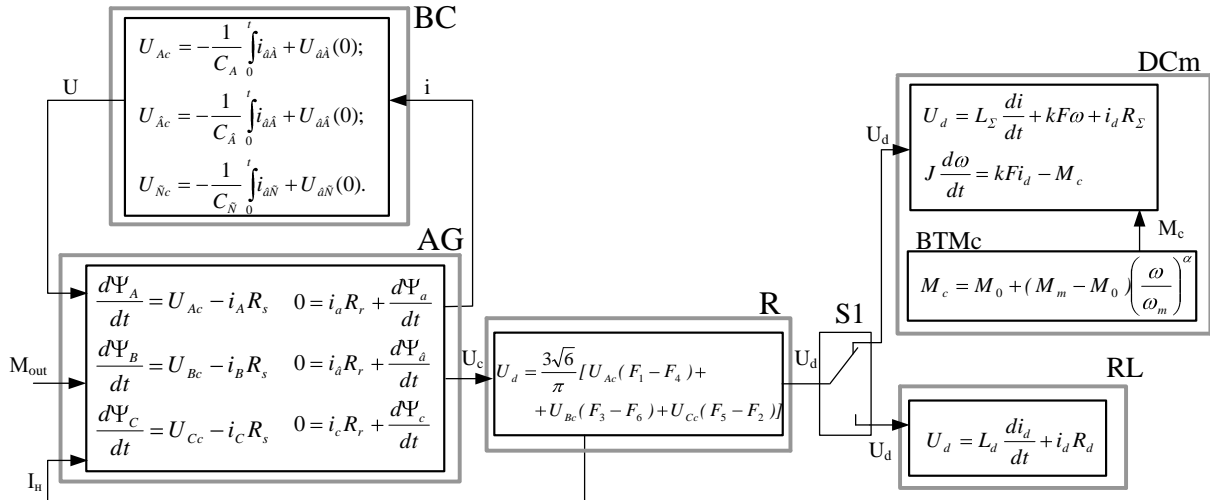


Fig. 3. Diagram of mathematical model of the AG-consumer system.

Transients of connecting and disconnecting of active, actively-inductive and motive loading are analyzed by using a computer modeling of the developed mathematical model. An asynchronous machine with the short circuited rotor of general industrial execution (type АИР80А4СУ2): PH=1,2 kW; nH=2740 rpm; I1=2,93 A; R1=9,37 Ohm; R2=5,13 Ohm; X1=7,03 Ohm; X2=6,5 Ohm, with capacitor excitation With=50 μF was used as a generator.

A power system operates on an extensive network and there could be few consumers on host load point. Their connection to supply chain is group mostly.

Dynamic characteristics of an asynchronous generator and two DC motors with independent excitation of different moments of resistance – DCm1 with the torque load of ventilator type ($M_c = f(\omega^2)$), DCm2 with the active torque load ($M_c = M_m = const$) are presented in fig. 4-6. DC power is PH_{DCm}=0,2 kW for each motor. Static error for voltage of an asynchronous generator is about 33% of original value,

dynamic – 68%. Through factor rectifier $k = \frac{3\sqrt{6}}{\pi}$ static error for voltage at the terminals of DC will be 9%.

DC starting could be considered as successful, but in the transition mode the generator is on the verge of the lower critical frequency of self-excitation. This can lead to loss of excitation of AG and the possibility of emergency.

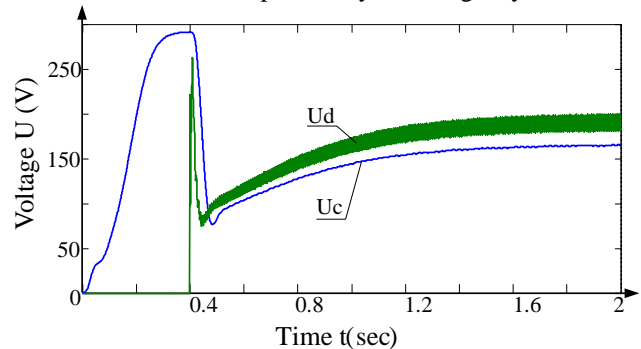


Fig. 4. Round generator voltage U_c and U_d rectified voltage at the terminals of DC.

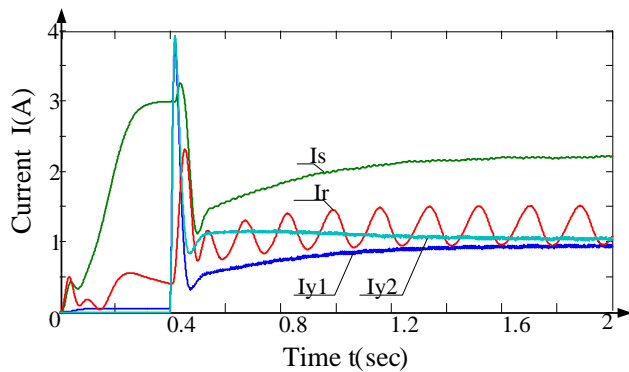


Fig. 5. Dependence of rounding currents the rotor I_r and stator I_s of AG, anchor chains DCm1 I_{y1} and DCm2 I_{y2} .

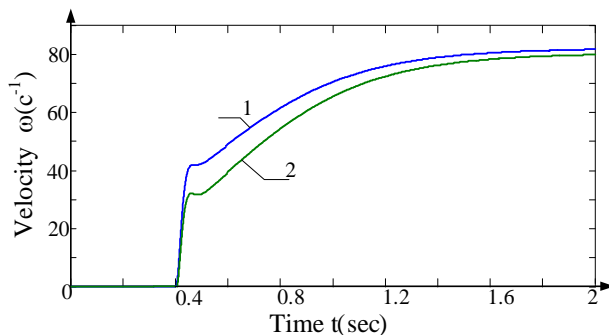


Fig. 6. Velocity of the DCm1 and DCm2.

In order to determine the ability of handling asynchronous generator with a capacitive excitation at work on DC consumers the relevant researches were done – step-by-step connection of different power DC motors (fig. 7).

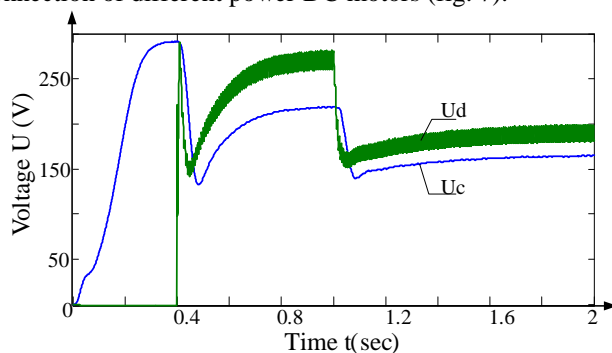


Fig. 7. Round generator voltage U_c and U_d rectified voltage at the terminals of DC.

Revealed that the successful start is possible when there is certain correlation of values for power sources and consumers:

$$\frac{P_{DCm}}{P_{AG}} \leq 0.35 \div 0.4. \quad (6)$$

Correlation (6) by 10-15% more than the relevant dependence while connecting of a single asynchronous motor like shown in [7].

Analysis of the research allows us to determine the specificity of physical phenomena in AG with capacitive excitation and with rectifier. In asynchronous generator with

capacitor excitation current of the capacitor (which contains a significant number of higher harmonic components) is magnetization current for generator. This causes distortion of the generated voltage. The above mentioned phenomenon is enhanced with high saturation of generator's magnetic core and cause to the not sinusoidal generated voltage. Switching diodes occurs earlier (raised the growing front and low falling) than it would take place at the sinusoidal voltage (fig. 8).

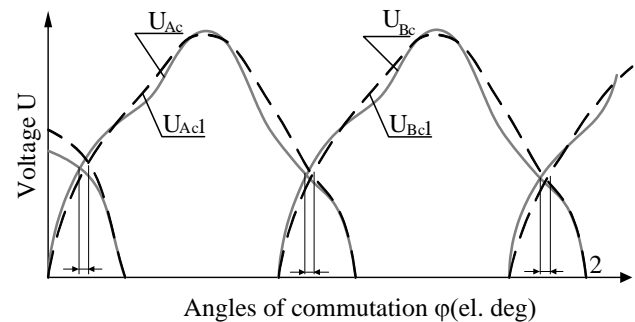


Fig. 8. Commutation processes in the system AG - uncontrolled rectifier.

IV. CONCLUSION

Research has shown (fig. 4, 7) that group start of electrical consumers is the most dangerous in terms of static and dynamic stability of the generator. This start is widespread in the industry. Therefore during design of ISE based on asynchronous generator, the above mentioned features should be considered while selecting the protective automation devices.

Results analysis showed that the uncontrolled rectifier output for the generator AG is an active-capacitive load and resulted to increase of rigidity for its external characteristics and handling ability.

REFERENCES

- [1] Вишневский Л.В. Системы управления асинхронными генераторными комплексами / Л.В. Вишневский, А.Е. Пасс. – К.: Либідь, 1990. – 168 с. (in Russian)
- [2] Гентковски З. Автономный асинхронный генератор с управляемым выходным выпрямителем. Тр. н.-т. конф. С международным участием, Крым, Алушта, 1995, с.229-232. (in Russian)
- [3] Зиновьев Г.С. Основы силовой электроники: Учеб. пособие. – Изд. 4-е, испр. и доп. – Новосибирск: Изд-во НГТУ, 2009. – 672 с. (in Russian)
- [4] Розанов Ю.К. Электронные устройства электромеханических систем: Учеб. пособие / Ю.К. Розанов, Е.М. Соколова – М.: «Академия», 2004. – 272 с. (in Russian)
- [5] Zacheпа Y. Research of the operation of the autonomous asynchronous generator with typical AC customers / XIII International PhD Workshop OWD 2011 - Wisla, 22-25 October 2011, pp. 413-418.
- [6] Чорний О.П. Моделирование электромеханических систем / Чорний О.П., Луговой А.В., Родькин Д.И. – Кременчук: 2001. – 376 с. (in Ukrainian)
- [7] Сыромятников И.А. Режимы работы асинхронных и синхронных двигателей / Под ред. Л.Г. Мамиконяца. – М.: Энергоатомиздат, 1984. – 528с. (in Russian)

Swing up of conductor depending on weight of ice and span size

(May 2012)

¹Matúš KATIN (3st year), ²Roman JAKUBČÁK (1nd year)
Supervisor: ³Ladislav VARGA

¹Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
^{2,3}Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

¹matus.katin@tuke.sk, ²roman.jakubcak@tuke.sk, ³ladislav.varga@tuke.sk

Abstract— The present article deals with dynamic phenomena on the external power line conductors, particularly with falling of ice (icing) from conductors and subsequent monitoring of amplitude of a conductor swing up. This paper compares the rate of impact of weight of ice and the span size on the size of swing up amplitude. The Simulation was performed with the use of a computer technique and the simulation software COSMOS – M.

Keywords—Icing, Swing up amplitude, Span size, Conductor

I. INTRODUCTION

Reliable operation of external power line conductors depends, inter alia, on climate conditions in which the power line is situated. In summer, when the weather is favorable, the power lines are loaded only with a static load. In winter, there is also a dynamic load added to the static load of the power line. A Higher dynamic load is caused by both wind and icing. Creation and subsequent falling of ice is the main cause of swinging up of conductor.

II. OBSERVATION OF IMPACT OF ICING WEIGHT TO THE SWING AMPLITUDE

The dependency of wire swing amplitude size when icing falling down and the weight of icing, which falls out from a given wire will be analyzed in the next part of the article.

Tab. 1 comparison of sizes of the wire swing AlFe 240/39 in the span of 300 meters in the middle if icing area as the dependency of weight of defecting icing

Weight of icing (kg/m)	sag in state 1 fl (m)	Mechanic stress in the state 1 (MPa)	Sag at the maximum amplitude of oscillation f _{mak} (m)	Wire swing fz (m)
0,5	10,596	55,01	9,962	0,635
1	10,944	71,20	9,714	1,230
1,5	11,265	86,60	9,473	1,792
1,7	11,395	92,97	9,372	2,024
2	11,564	101,33	9,237	2,327
2,5	11,844	115,49	9,007	2,837
3	12,109	129,17	8,783	3,327
4	12,600	155,27	8,349	4,251
5	13,048	179,99	7,934	5,114
6	13,462	203,57	7,534	5,928
7	13,848	226,20	7,148	6,700
8	14,211	248,00	6,773	7,438
9	14,553	269,09	6,405	8,148
10	14,878	289,53	6,044	8,833

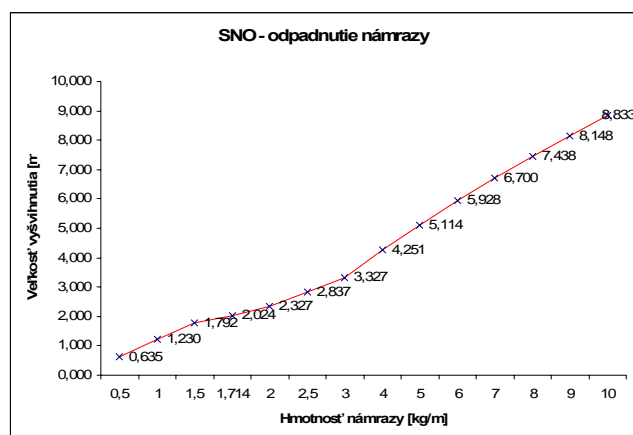


Fig. 1 graphic dependency of wire swing size AlFe 240/39 in the span of 300 meters in the middle of the icing area in the dependency of weight of defecting icing

From the given facts we can deduce that the size of icing, more exactly its weight, there is a significant impact to the size of wire swing amplitude. When simulating the icing 3 kg/m and its consecutive defection there was an incrementing wire swing amplitude. In the case of icing 4 kg/m there is a higher probability of dangerous approach of neighboring wires and consecutive short circuit between phases.

In the process of icing defection simulation it is significant to realize that the ideal case of icing defection from the wire at once in the direction perpendicular to the earth surface there are hypothetical oscillations of wire in the vertical direction. At this ideal move there is an assumption that the normal wire in the keg arrangement of the double air wiring is moved at about of 0.8 meter towards the wires placed above and below it, there an is assumption that the wire would not touch the wire above it. At most the wires could get close to each other at the distance that equals the difference from their coordinates, thus for example 0.8 m. however this ideal happening of wire in the dynamic process of icing defection at the ideal climatic conditions without the wind occurrence has a low probability.

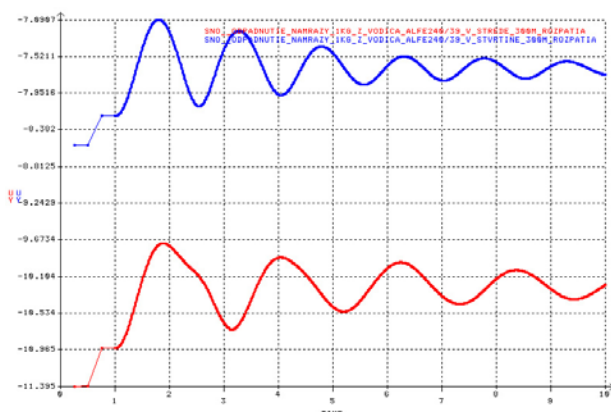


Fig. 2 Time dependency of wire move AlFe 240/39 measured in the middle and a quarter of span of 300 m at the time of 1 kg/m icing defection at design for a middle icing area.

From the data in the table is consequent that at the icings of unitary weight higher than 10 kg/m there would appear a

mechanic tension in the wire, with higher values than its measuring firm, which is in this case 273.18 MPa by the manufacturer data

III. THE IMPACT OF SPAN SIZE TO THE CONDUCTOR SWING

Tab. 2 dependency of wire swing to the wire span at watching the icing defection from the wire AlFe 240/39 placed in the middle of icing area

Span	Sag in the state 0	Sag at the maximum amplitude of oscillation	Wire swing
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 2 are shown maximum amplitudes of swing which the wire raised at the simulation of icing falling down. In the figure 3 is shown the dependency of wire swing amplitude at falling of icing to the span size.

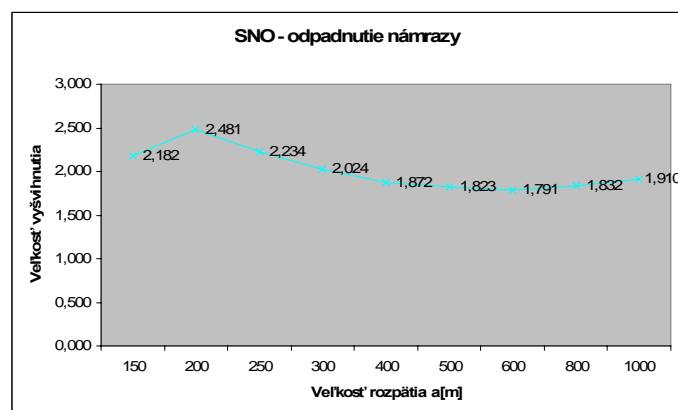


Fig. 3 graphic dependency of wire swing size to the size of span at watching the icing defection from wire AlFe 240/39 placed in the middle of icing area

Following waveforms show the time progress at wire swing at the time of icing defection observed in various distances of the span.

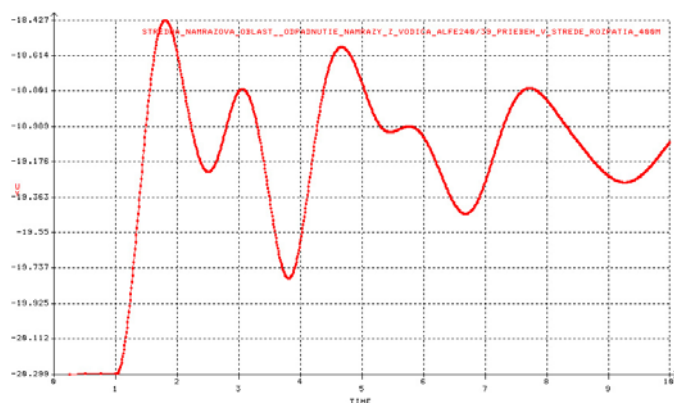


Fig. 4 time dependency observed in the middle of 400 m span at icing defection from the wire AlFe 240/39 placed in the middle of icing area

IV. CONCLUSION

As the conclusion after evaluation of all the simulations of wire swing we can deduce the following conclusions:

- The weight of icing on the wire has a significant impact on the dynamic phenomenon as well as on the static weight load of the wire. As it has been mentioned and as it is apparent from the table number 1 the herniation of the wire can occur when there is a higher load of icing to the wire
- Table number 1 and the graphic dependency in the picture number 1 show the fact that the weight has a significant impact to the amplitude of wire swing. At simulations there were observer amplitudes which reached the values of 8 meters. At these amplitudes we can observe that the touch of dangerous approach of wires can occur
- The executed simulations of impact of the icing area and span distance show the fact that both the factors do not have a significant impact to the amplitude of wire swing as the weight of the icing has

REFERENCES

- [1] Hudák, R.: Riešenie mechaniky kotevného poľa vonkajšieho silového vedenia (VSV) v priestore. Dizertačná práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2010. 130 s.
- [2] Fecko, Š.: Elektrické siete. Vonkajšie vedenia. SVŠT, Elektrotechnická fakulta, Bratislava, 1978.
- [3] Leščinský, P., Varga, L., Segľa, Š., Ilenin, S.: Modelovanie pohybu vodičov vzdušných silových vedení pri opadávaní námrazy. In: Elektrotechnika v praxi, 2005, č. 2, str. 107 – 109.

The development of experimental equipment for investigation the induction motors diagnostic system based on the electrical signals analysis

¹Dmytro MAMCHUR, ²Andrii KALINOV, ³Tibor VINCE

^{1,2} Institute of Electromechanics, Energy Saving and Control Systems,
Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine

³Dept. of Theoretical Electrotechnics and Electrical Measurement, FEI TU of Košice, Slovak Republic

^{1,2}scenter@kdu.edu.ua ³tibor.vince@tuke.sk

Abstract – In the article the developed hardware and software for induction motor diagnostic system is presented. Hardware allows data measurement for different operating modes with different loading of motors with most common damage types. Software allows make analysis of preliminary recorded data using motor current signature analysis and instantaneous power spectra analysis methods. Experimental tests with real data analysis showed possibility utilize developed equipment for induction motor diagnostic systems.

Keywords – Diagnostics, induction motor, instantaneous power, spectra analysis.

I. INTRODUCTION

Induction machines (IM) play an important role in industry that's why there is exacting requirements for their reliable and safe operation. Faults and failures of induction machines can generate losses in terms of maintenance and lost revenues, and also lead to excessive downtimes. All these facts are motives of on-line condition monitoring. Nowadays there are some well-known on-line monitoring methods [1]. The analysis of these methods had shown that instantaneous power analysis as a medium for on-line fault diagnostic of IM is a very attractive for analysis of low-power machines. This method does not need expensive equipment, but need just currents and voltage sensors. As for analysis, it had to be mentioned, that the amount of information is carried by instantaneous power. And in difference to motor current signature analysis (MSCA) the instantaneous power spectra, besides the fundamental and the two classical sideband components, contains additional components directly at the modulation frequency [2]. All these facts described necessity of creation laboratory equipment for IM diagnostic systems based on instantaneous power analysis.

II. HARDWARE DEVELOPMENT

The most convenient IM diagnostic methods are the following:

1) supply voltage analysis;

2) motor currents signature analysis (MCSA);

3) instantaneous power spectra analysis.

The result of supply voltage analysis essentially depends on power supply quality, thus this method is unsuitable for low-voltage networks with polyharmonic voltage.

Motor stator currents analysis is based on supervision of change the air gap between the stator and rotor which is back reflected in the form of the motor current through change of a magnetic flux in the air gap, which causes counter-emf. These changes in a counter-emf eventually change a wave of IM current. Fast Fourier Transform (FFT) of motor current gives a current spectrum for fault detection procedure.

Thus, presence in a motor current spectrum specific harmonics and their sidebands shows presence of electrical or mechanical damages.

This method could be used for diagnostic the following defects:

- rotor unbalance;
- broken rotor bars;
- bearings damage.

This method also depends on power supply quality, and in some cases could lead to wrong results.

Lacks of the previous methods can be avoided using the diagnostics on the basis of the instantaneous power spectra analysis [2, 3]. Instantaneous power spectra analysis allows both detection of fault presence and estimation of damage level by analysis of proper harmonic value. Thus, it allows one to make estimation of the energy of fault and the correlation of this energy to additional damage of IM parts under influence of additional vibrations caused by proper harmonic. Moreover, the instantaneous power spectra analysis allows analyzing of IM operation modes under significant nonlinearity, when it is incorrect to use superposition principle for current harmonics. Also, instantaneous power analysis is more reliable, it is less dependent on noise, and gives additional harmonic components for analysis [2].

In work [4] the expressions for calculation total 3-phase instantaneous power spectra harmonics related to most frequently caused damages were given.

III. HARDWARE DEVELOPMENT

For imitation the most frequently caused defects of IM, such as rotor bar breaks, stator windings asymmetry and short circuits, bad mounting to the ground and rotor eccentricity, the three identical induction motors of type AIP80B4Y2 1.5 kW were used. Their design provides the possibility of creation the rotor bar break and turn-to-turn short circuit (Fig. 1, 2).

In order to imitate turn-to-turn short circuits, the taps were provided in one of the phases of stator winding (Fig. 1, Table 1). To imitate rotor bar breaks, apertures were drilled in the places where bars attach cage rings, with the aim of breaking the contact of the bar and the ring. For broken bars investigation several rotors of identical type with 1, 2 or 3 broken bars, which can be interchanged, were used. A scheme of apertures location is shown in Fig. 2.

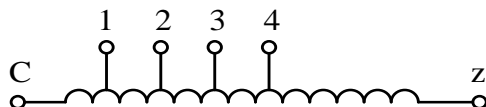


Fig. 1. Stator phase winding taps circuit.

TABLE I

IM WINDING RESISTANCE MEASUREMENT DATA

Phase	Resistance value, ohm		
A	5.102		
B	4.972		
C	4.945		
Taps in phase C	Winding part	Resistance value, ohm	Reduction of winding turns number, %
	1-z	4.93	0.3
	2-z	4.878	1.36
	3-z	4.82	2.52
	4-z	4.253	14

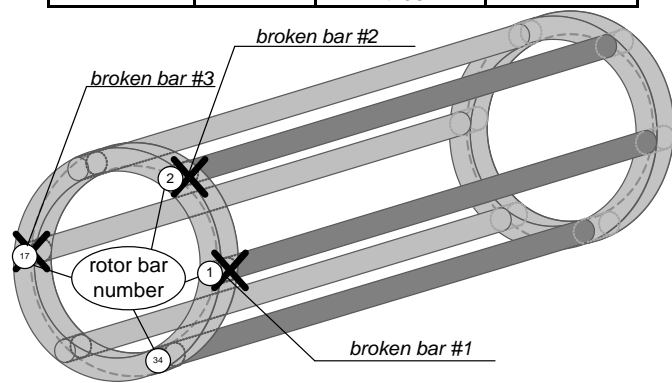


Fig. 2. Scheme of rotor apertures location

DC generator provided a mechanical load. Motor angular speed could be controlled by speed sensor. Supply unsymmetry could be done by using voltage transformer in one of the supply phases. Measuring modules with measuring software was developed by authors for measurement and record the electrical values (voltages and currents) which are necessary for the analysis (Fig. 3). The photo of created experimental equipment is shown in Fig. 4.

IV. SOFTWARE DEVELOPMENT

The software for IM fault detection system was developed basing on the expressions for correspondence of fault and electrical signal harmonic component known from MSCA

analysis method [1] and the instantaneous power analysis method [4]. The user interface is shown in fig. 5–8. This software allows us to upload preliminary measured data for analysis (Fig. 5). Also there is possibility define analyzed signal length. Basing on given signals, program calculates FFT transform and shows signals spectra for each phase electric signals (Fig. 6) and total three phase instantaneous power signal (Fig. 7). According to developed algorithms, program gives diagnostic result based both on MSCA method and instantaneous power analysis method (Fig. 8).

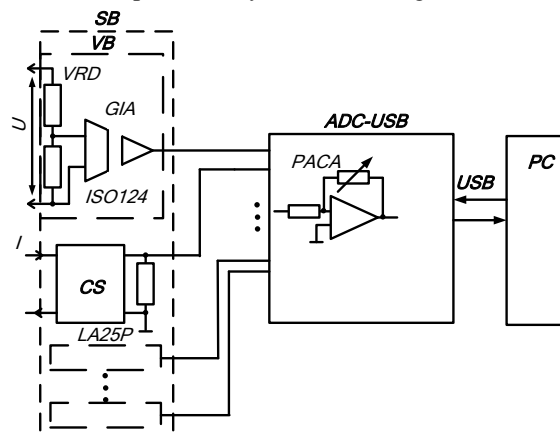


Fig. 3. Measuring module functional circuit:

SB – sensor block;
VB – voltage block;
VRD – voltage resistance divider;
GIA – galvanic isolation amplifier;
CS – current sensor;
PC – personal computer;
PACA – programmed amplification coefficient amplifier;
USB – PC bus.

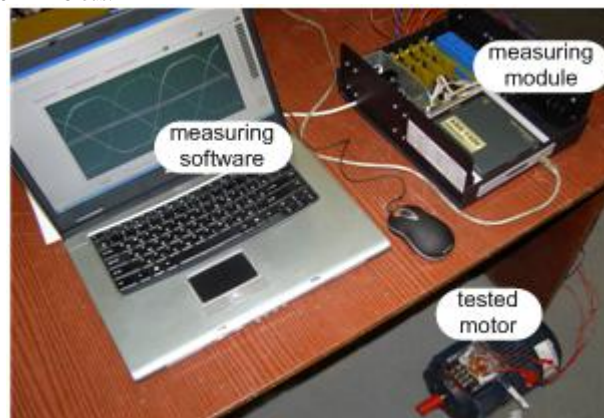


Fig. 4. Photo of the measuring complex

V. EXPERIMENTAL RESEARCHES

Analysis was carried out for IM idle conditions and for full load conditions. Preliminary data showed that defects could be more obviously manifested in loaded machine signals.

The analysis of the fulfilled experiments makes it possible to come to the following conclusions. When rotor bar break is imitated and motor works with full load, a growth of sideband harmonics at frequencies of $50n \pm 2.5 \text{ Hz}$ ($n=1,2,3,\dots$) in the current spectrum can be observed. Meanwhile, three phases total power signal besides sideband harmonics contain low-frequency component at frequency of 2.5 Hz, which is additional diagnostic feature. Due to cross-connections during the formation of instantaneous power harmonics components

on the basis of current and voltage components, this component reveals in the power signal at the frequencies divisible by 2.5 Hz. This component can be hardly noticeable, because of its small amplitude, at the frequencies divisible by

the frequency of supply mains (50 Hz). Thus, it is possible to determine the existence of rotor bar breaks by amplitude values of harmonics of $50n \pm 25$ Hz.

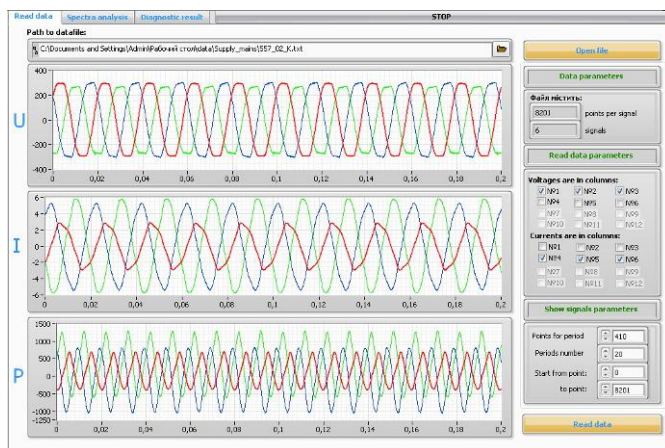


Fig. 5. Read data and signal settings window

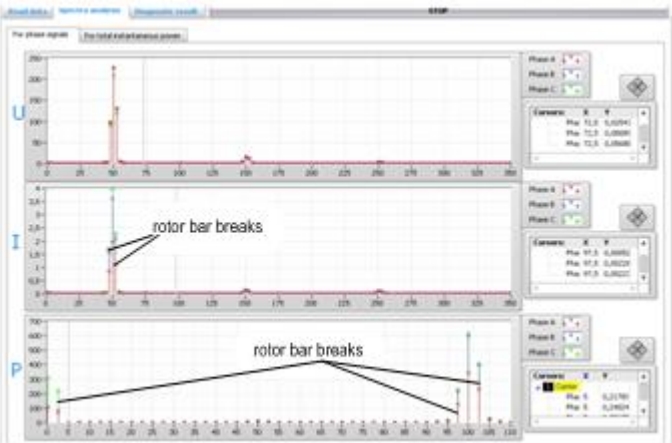


Fig. 6. Spectra analysis for phase electric signals

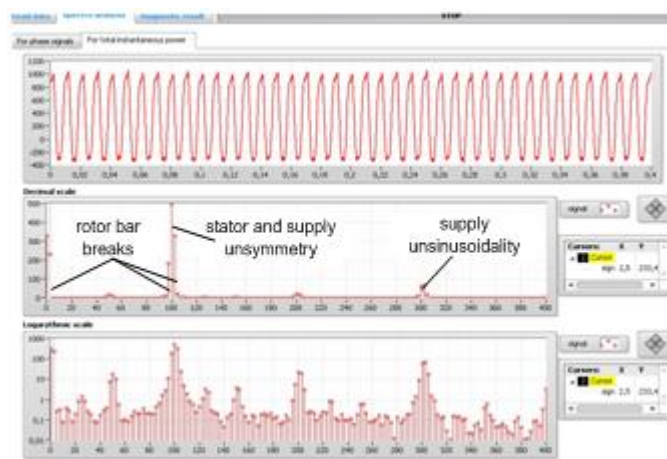


Fig. 7. Spectra analysis for total 3-phase instantaneous power signal



Fig. 8. Diagnostic result window

Growth of the harmonic amplitude in total 3-phase power at frequency of 300 Hz can be explained by the power supply uninusoidality.

When turn-to-turn short circuits appear, as well as when there is a stator windings asymmetry, the amplitude of harmonics at frequency of 100 Hz in phases power and total 3-phase power increases significantly. This feature also absent for MSCA analysis.

Preliminary analysis for investigated motor provided using MSCA and Instantaneous power analysis techniques showed, that under full load conditions both methods showed perfect results for broken bars detection. But for idle mode broken bars were detected just using Instantaneous power spectra analysis (Fig.8). The same results were for rotor unbalance detection under idle conditions.

Thus, preliminary researches lead to conclusion that both methods are reliable to detection most common IM faults for loaded motor, but in case of idle mode the Instantaneous power spectra analysis gives much better results because of additional harmonic components related to damages.

VI. CONCLUSION

Review of modern methods for IM fault detection resulted in conclusion of necessity development testing equipment and software for simplifying diagnostic operation. Most easy-used and reliable methods among reviewed are Motor current signature analysis and Instantaneous power spectra analysis. Thus, developed software provides to make analysis based on both these methods. Experimental tests with real data analysis showed possibility utilize developed software for IM diagnostic systems and showed advantages of Instantaneous power analysis for idle mode.

REFERENCES

- [1] M. E. H. Benbouzid "A review of induction motors signature analysis as a medium for faults detection", IEEE Transactions on Industrial Electronics, vol. 47, no. 5, pp. 984–993, Oct. 2000.
- [2] S.F. Legowski, A.H.M. Sadrul Ula and A.M. Trzynadlowski, "Instantaneous power as a medium for the signature analysis of induction motors", IEEE Transactions on Industrial Electronics, vol. 32, no. 4, pp. 904–909, Jul./Aug. 1996.
- [3] A.M. Trzynadlowski and E. Ritchie, "Comparative investigation of diagnostic media for induction motors: a case of rotor cage faults", IEEE Transactions on Industrial Electronics, vol. 47, no. 5, pp. 1092–1099, Oct. 2000.

- [4] D. Mamchur “An instantaneous power spectra analysis as a method for induction motors fault detection”, Proceedings of OWD’2011, 22-25 October 2011, Wisla, pp. 407-412, ISBN 83-922242-4-0

The diffusion coefficient for nonlinear Brownian motion and the electric circuits

¹Lukáš GLOD, ²Gabriela VASZIOVÁ

¹Dept. of Mathematics and Physics, Institute of Humanitarian and Technological Sciences,
The University of Security Management, Košice, Slovak Republic

²Dept. of Physics, FEI TU of Košice, Slovak Republic

¹lukas.glod@vsbm.sk, ²gabriela.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 this paper one example of such generator is studied. The dissipative friction force is taken from the current models of Brownian motors and is nonlinear. We have also obtained an exact expression for the effective diffusion coefficient, if the Langevin equation includes the spurious force. This result can be used to find the values of system parameters m , T and γ for many other models. We show that the diffusion coefficient of nonlinear Brownian motion is inversely proportional to the friction coefficient γ .

Keywords—Diffusion coefficient, Langevin equation, nonlinear friction, noise generator.

I. INTRODUCTION

In the article published in Comptes Rendus in 1908 Paul Langevin proposed a different approach to the 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].

According to the classical theory the motion of the BP with the mass m and radius R in a fluid is described by the Langevin equations (d/dt is denoted by a dot)

$$\dot{x} = v, \quad m\dot{v} = -\gamma v + \sqrt{2D}\xi(t), \quad (1)$$

where $v(t)$ is the particle velocity and $x(t)$ is the particle position. The resistance force during the particle motion is the Stokes friction force proportional to the velocity, and the irregular impulses from the surrounding molecules are described by the (white) noise force $\sim \xi(t)$ with the properties

$$\langle \xi(t) \rangle = 0, \quad \langle \xi_i(t) \xi_j(t') \rangle = \delta_{ij} \delta(t - t'). \quad (2)$$

In (1), $D = k_B T \gamma$ is the noise intensity (k_B is the Boltzmann constant, T is the temperature and $\gamma = 6\pi R \eta$ the friction coefficient proportional to the dynamic viscosity of the

surrounding fluid η). The effective diffusion coefficient of the particle, which also determines the intensity of the noise generated by (1), is given by the Einstein formula

$$D_{\text{eff}} = \gamma^{-2} D. \quad (3)$$

Equation (1) can be easily solved with the properties (2). The derivation of the mean square displacement and the velocity autocorrelation function from (1) can be found in many textbooks and monographs [3].

Various generalizations of (1) during the past 20 years were used to describe the motion of living objects – molecular motors, flocks of animals, etc. The simplest case for the friction force of the individual BP 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 acts as an energy pump. This Rayleigh-Helmholtz model is a standard model studied in many papers on Brownian dynamics [4 - 7]. We note that v_0^2 defines a special value of the velocities where the friction is zero. Another models (not addressed here) that have attracted some attention were proposed by Schweitzer, Ebeling and Tilch in [8], and by Schienbein and Gruler in [9].

In this paper we study the one-dimensional case when the friction coefficient γ is replaced by the symmetric function of velocity

$$\gamma(v) = \gamma v^{2\alpha}, \quad (4)$$

where α is a constant.

Lindner [10] worked out in detail the dependence of the effective diffusion coefficient on the system parameters D , γ , and m for the BM characterized by a friction force (4), if $\alpha = 0, 1, 2, 3$, and the noise intensity D stayed a constant. Generally, the noise intensity is no more constant for BM in which the friction function depends on the particle’s velocity in a nonlinear way.

Our main focus is to derive the effective diffusion coefficient for a nonlinear BM with a general friction function. In this case the equation of motion for the BP should contain a stochastic (or “spurious”) force proportional to $\partial_v D(v)$ [4]. Whereas in [11] the exact expression for the diffusion coefficient of the nonlinear BM when the Langevin equation does not include the spurious force was derived, here we take this force into account.

II. THE DIFFUSION COEFFICIENT IN THE CASE OF NONLINEAR FRICTION

In the mentioned work [10], the main attention was given to the effective diffusion coefficient of the particle D_{eff} . This coefficient can be calculated from the mean square displacement of the particle as

$$D_{\text{eff}} = \lim_{t \rightarrow \infty} \frac{1}{2dt} \sum_{j=1}^d \left\langle [x_j(t) - x_j(0)]^2 \right\rangle, \quad (5)$$

or using the Kubo relation for the velocity autocorrelation function,

$$D_{\text{eff}} = \frac{1}{d} \int_0^\infty d\tau \langle v(t)v(t+\tau) \rangle. \quad (6)$$

Here, d is the dimensionality of the system. The coefficient D_{eff} is important not only for the description of various kinds of the active and nonlinear BM, e.g. in physics and biology [12 - 14], but also for designing colored-noise generators [10]. 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 equation (1) with nonlinear frictions (4) 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 quantity and its dependence on the system parameters D , γ , and m .

Lindner’s theory requires substantial improvements since for the nonlinear BM it incorporates the same random force as for the linear case. In the correct approach, when the friction force nonlinearly depends on the velocity v , the random force should be changed: the intensity of the noise must also be the function of v , $D(v)$, as an implication of the fluctuation-dissipation theorem. Due to this fact we have to add “stochastic” or the so called spurious force, proportional to the derivative of $D(v)$ [4] in the Langevin equation. Equation (1) then becomes

$$m\dot{v} = -\gamma(v)v - am^{-1}\partial_v D(v) + \sqrt{2D(v)}\xi(t), \quad (7)$$

where the moments of random function $\xi(t)$ are defined as before. Equivalent to the Langevin equation (7) is the Fokker-Planck equation for the distribution function $P(v,t)$

$$\partial_t P = m^{-2}\partial_v [D(v)\partial_v P] + m^{-1}\partial_v \left\{ \left[\gamma(v)v + m^{-1}(a+1/2)\partial_v D(v) \right] P \right\}. \quad (8)$$

The equilibrium solution of the Fokker-Planck equation (8) must have the form of the Maxwell-Boltzmann distribution. After substituting this distribution into (8) for the state of equilibrium, the following equation for the case in question the fluctuation-dissipation relation

$$D(v) = k_B T \left[\gamma(v) + m^{-1}(a+1/2)v^{-1}\partial_v D(v) \right] \quad (9)$$

must reduce to the generalized Einstein relation $D(v) = k_B T \gamma(v)$, and hence the coefficient a in the equations of Langevin and Fokker-Planck must be $a = -1/2$.

Equation (8) can be established in different ways. Its most typical forms are: Ito’s [15, 16], Stratonovich’s [17, 18], and the kinetic’s form [4, 19, 20], corresponding to $a = 1/2, 0, -1/2$, respectively, and it is only in the last case that (9) coincides with the generalized Einstein relation [4].

In the kinetic interpretation, the Fokker-Planck equation has

the stationary solution $P_{\text{st}}(v) = (m/2\pi k_B T)^{1/2} \exp(-mv^2/2k_B T)$. The normalization constant was determined from the normalization condition $\int_{v_{\min}}^{v_{\max}} dv P_{\text{st}}(v) = 1$ for the natural boundary conditions ($v_{\min} = -\infty, v_{\max} = +\infty$).

Now, we obtain the general formula for the effective diffusion coefficient D_{eff} for the one-variable case corresponding to kinetic Fokker-Planck equation (8). According to [14, 21], the diffusion coefficient is given by the simple formula

$$D_{\text{eff}} = \langle \Delta v^2 \rangle \tau_{\text{corr}}, \quad (10)$$

where $\tau_{\text{corr}} = \int_0^\infty d\tau [\langle v(t)v(t+\tau) \rangle - \langle v \rangle^2] / \langle \Delta v^2 \rangle$ is the correlation time of the velocity and $\langle \Delta v^2 \rangle = \langle (v - \langle v \rangle)^2 \rangle$ is its variance. On the other side, in Ref. [22] we can find the analytical expression for the correlation time (S9.14). Using this fact, we can obtain the exact formula for the effective diffusion coefficient D_{eff} in the one-dimensional case ($d = 1$),

$$D_{\text{eff}} = \sqrt{\frac{mk_B T}{2\pi}} \int_{v_{\min}}^{v_{\max}} dv \gamma^{-1}(v) e^{-U(v)}, \quad (11)$$

where $U(v) = mv^2/2k_B T$. If the general friction function $\gamma(v)$ is symmetric in v , e.g. (4), and “well-behaved” avoiding unphysical divergence of the velocity, in particular, the velocity can vary between two symmetric limiting values $\pm v_M$ [11] and the formula (11) can be replaced by

$$D_{\text{eff}} = 2k_B T \int_0^\infty dv \gamma^{-1}(v) P_{\text{st}}(v). \quad (12)$$

The result (11) is defined for kinetic representation of the Fokker-Planck equation, which corresponds to the Langevin equation (7) with the spurious force. It is easily verified from (11) that the diffusion coefficient is proportional to a simple power of the friction coefficient γ like γ^{-1} . For normal BM ($\alpha = 0$) the relation (3) takes place.

The diffusion coefficient in our nontrivial example, which is a nonlinear function (4) discussed in [10], reads

$$D_{\text{eff}} = \frac{2^{-\alpha} (k_B T)^{1-\alpha} m^\alpha}{\sqrt{\pi} \gamma} \Gamma\left(\frac{1}{2} - \alpha\right), \quad (13)$$

if $2\alpha < 1$. This result coincides with the result from dimensional analysis. For $\alpha = -1$ from (13) one recovers the simple relation $D_{\text{eff}} = k_B^2 T^2 / m\gamma$.

III. THE DIFFUSION COEFFICIENT FOR ELECTRIC CIRCUIT

In this final part of our contribution we will show that the presented theory can be equally applied to quite a different kind of problem, namely, for electric circuits. It is well known that e.g. fluctuations in electric circuit that are in contact with a thermal bath and BM of particles can be described by essentially the same mathematics, the standard Langevin equation of motion [23, 24]. Equation (7) exactly corresponds to a simple electric circuit in which the resistor with the resistance R and the inductor with inductance L are connected in series. They are subject to voltage source $V(t)$. There is also a thermal noise generator next to the resistance, which reflects the fluctuations of the voltage drop across the resistor, $\delta V(t)$. The equation for such a circuit is

$$L\dot{I} + R(I)I - \frac{k_B T}{2L} \partial_I R(I) = \sqrt{2k_B T R(I)} \delta V(t). \quad (14)$$

The correspondence between (7) and (14) is seen if the particle velocity v is replaced by the current I , L replaces the particle mass m , and R is for the friction coefficient γ . Energy is being dissipated in the resistor. The fluctuations that accompany this dissipation are described by a random noise term $\delta V(t)$. Using the described correspondence between the circuits and the nonlinear BM, the solution for the diffusion coefficient is directly given by the expressions in the previous section.

It is important that at the frequency $\omega = 0$ the spectral density of the current, $(I^2)_{\omega}$, can be regarded as the diffusion coefficient D_{eff} (11).

IV. CONCLUSION

In this contribution, we have studied the famous Langevin equation as a generator of the colored noise. We have described an 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 [10]. The resulting Langevin equation is nonlinear and hardly solvable analytically. However, it is possible to obtain an exact expression for the effective diffusion coefficient.

In Section II, we have derived an analytical expression for the diffusion coefficient of a one-dimensional Brownian motion with nonlinear friction, taking into account the spurious force in the Langevin equation. This coefficient is inversely proportional to the friction factor γ . We have shown that the expected Einstein formula takes place (3) only in the case when the friction force $\gamma(v) \rightarrow \gamma$ (normal Brownian motion). This expression exactly corresponds to the theory. Expression (11) can be also used to explore other issues. Currently, our attempts are oriented in this direction.

At present the theory of the nonlinear Brownian motion is intensively developed. We believe that along with the remarkable improvements of the experimental possibilities and broadening of the observable time and space scales it will find more and more applications, particularly in the science and technology of electrical engineering.

ACKNOWLEDGMENT

We thank our supervisor, Prof. V. Lisý, for introducing us to the theme and for continuous support.

This work was supported by the Agency for the Structural Funds of the EU within the NFP project No. 26220120033, and by the grant VEGA 1/0370/12.

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] P. Langevin, "Sur la théorie du mouvement brownien", *C. R. Acad. Sci.* 146, 530 (1908) (English translation: A. Gythiel, "On the theory of Brownian motion", *Am. J. Phys.* 65, 1079 (1997)).
- [4] Yu. L. Klimontovich, "Nonlinear Brownian motion", *Phys. Usp.* 37, 737 (1994).
- [5] U. Erdmann, W. Ebeling, L. Schimansky-Geier, A. Ordemann, F. Moss, zooplankton: Application to experiments with swarms of *Daphnia*", arXiv:q-bio.PE/0404018 (2004).
- [6] U. Erdmann, W. Ebeling, L. Schimansky-Geier, F. Schweitzer, "Brownian particles far from equilibrium", *Eur. Phys. J. B* 15, 105 (2000).
- [7] U. Erdmann, W. Ebeling, V. S. Anishchenko, "Excitation of rotational modes in two-dimensional systems of driven Brownian particles", *Phys. Rev. E* 65, 061106 (2002).
- [8] F. Schweitzer, W. Ebeling, B. Tilch, "Complex motion of Brownian particles with energy depots", *Phys. Rev. Lett.* 80, 5044 (1998).
- [9] M. Schienbein, H. Gruler, "Langevin equation Fokker-Planck equation and cell migration", *Bull. Math. Biol.* 55, 585 (1993).
- [10] B. Lindner, "Diffusion coefficient of a Brownian particle with a friction function given by a power law", *J. Stat. Phys.* 130, 523 (2008).
- [11] B. Lindner, "The diffusion coefficient of nonlinear Brownian motion", *New J. Phys.* 9, 136 (2007).
- [12] P. Reimann, "Brownian Motors: Noisy Transport far from Equilibrium", *Phys. Rep.* 361, 57 (2002).
- [13] W. Ebeling, L. Schimansky-Geier, "Swarm dynamics, attractors and bifurcations of active Brownian motion", *Eur. Phys. J. Special Topics* 157, 17 (2008).
- [14] B. Lindner, E. M. Nicola, "Diffusion in different models of active Brownian motion", *Eur. Phys. J. Special Topics* 157, 43 (2008).
- [15] K. Ito, "Stochastic integral", *Proc. Imp. Acad.* 20, 519 (1944).
- [16] K. Ito, "On stochastic differential equations", *Mem. Am. Math. Soc.* 4, 51 (1951).
- [17] R. L. Stratonovich, "A new representation for stochastic integrals and equations", *Vestnik Moskov Univ. Ser. 1 Mat. Meh.* 1, 3 (1964).
- [18] R. L. Stratonovich, "A new representation for stochastic integrals and equations", *SIAM J. Control* 4, 362 (1966).
- [19] P. Hänggi, "Stochastic processes I: asymptotic behaviour and symmetries", *Helv. Phys. Acta* 51, 183 (1978).
- [20] P. Hänggi, "Connection between deterministic and stochastic descriptions of nonlinear systems", *Helv. Phys. Acta* 53, 491 (1980).
- [21] B. Lindner, "Diffusion of particles subject to nonlinear friction and a colored noise", *New J. Phys.* 12, 063026 (2010).
- [22] H. Risken, *The Fokker-Planck Equation*. Berlin: Springer, 1984.
- [23] R. van Zon, S. Ciliberto, E. G. D. Cohen, "Power and heat fluctuation theorems for electric circuit", *Phys. Rev. Lett.* 92, 130601 (2004).
- [24] N. Garnier, S. Ciliberto, "Nonequilibrium Fluctuations in a Resistor", *Phys. Rev. E* 71, 060101(R) (2005).

The energy method for identification the electromagnetic parameters of induction motors

¹Yuriy ROMASHIHIN, ²Martin BAČKO (3rd year), ³Ján MOLNÁR

Supervisor: ¹Dmitriy RODKIN

¹ Institute of Electromechanics, Energy Saving and Control Systems,
Kremenchuk Mykhailo Ostrohradskyi National University, Ukraine

^{2,3}Dept. of Theoretical Electrotechnics and Electrical Measurement, FEI TU of Košice, Slovak Republic

¹Romashihin_yuriy@mail.ru, ²martin.backo@tuke.sk ³jan.molnar@tuke.sk

Abstract—In this paper the energy method for identification the electromagnetic parameters of induction motors is described. The method is based on the use of the balance equations for the instantaneous power components.

Keywords—electromagnetic parameters, induction motor, instantaneous power.

I. INTRODUCTION

To date, the induction motors (IM) with a squirrel-cage rotor are most widely used for electric drive systems. They are characterized by low cost, high reliability and satisfactory performance. But due to various factors, the high reliability of IM for nominal conditions does not coincide with the real. This leads to the choice of engine over-capacity, an increase in material costs for repairs and service. Analysis of Ukrainian enterprises showed that the average yield of IM in the repair approximately 20-50% [1, 2].

The reasons for failure of IM are the lack of the service, poor quality of repair, the natural aging process, violations of the production technology. This leads to a change in the electromagnetic parameters. Change the electromagnetic parameters cause additional errors in the calculation of starting currents and torques, losses in the steel and copper, configuring systems of the motor control.

Therefore there is a need to search for and evaluation method for the determination of IM electromagnetic parameters. Analysis of existing methods for the determination of electromagnetic parameters of IM was showed that the majority of them is low-tech and does not satisfy the requirements of the practical problems of the modern drive. Known methods for the determination of electromagnetic parameters require a partial or complete disassembly of the engine, install additional equipment on the engine components, and have a low accuracy or low level of automation. Almost all methods allow you to define only a portion of the electromagnetic parameters of IM.

Analysis [2, 3] was showed that at the present time are the common methods of identifying the frequency of the electromagnetic parameters of IM. Most of them are based on the analysis of transient process. As a separate class of

methods can be viewed using the polyharmonic power of the IM. Among them are the energy method [2].

II. THE OBJECT OF THE PAPER

Estimate of efficiency of the energy method for the identification of electromagnetic parameters of induction motors.

III. MATERIAL AND RESULTS OF THE RESEARCH

Analysis of energy process in the IM should be carried out using the power balance equations. Effective application of this approach is possible with the set of equations for the components of the instant power at all elements of the «power source – an induction motor». The basic balance equation is the equality of the instant power of source $p_s(t)$ and instant powers sums of elementary consumers in the system [1, 2]:

$$p_s(t) = \sum_{i=1}^H p_i(t), \quad (1)$$

where: i – the index of the corresponding elementary consumer; H – the number of elementary consumers.

The balance of power is the result of theorems Telldzhen: the sum of products of currents I_i and voltages U_i all branches of the chain, satisfying Kirchhoff's laws, is equal to zero. The product is $U_i I_i$ instant power P_i branch, so the sum of the capacities of all branches of the circuit is zero. If you select a branch from independent sources, the balance of power can be summarized as follows: the sum capacity given by independent sources equals the amount of power consumed by the other branches of the circuit.

With regard to the well-known T-equivalent circuit of IM as basic consumer components, each of the parameters of the equivalent circuit (fig. 1). This inductance of stator L_1 and rotor L_2 , inductance of the magnetizing circuit L_μ , the resistance of the stator R_1 , rotor resistance R_2 and the resistance of the circuit of the magnetization R_μ . The instant power balance equations for single-phase motor shall be made in the form of equations constituting the instant power sources and instant power on all elements of the equivalent circuit [2, 3]:

$$p_s(t) = p_{R_1}(t) + p_{R_2}(t) + p_{R_\mu}(t) + p_{L_1}(t) + p_{L_2}(t) + p_{L_\mu}(t), \quad (2)$$

where: $p_{R_1}(t)$, $p_{R_2}(t)$, $p_{R_\mu}(t)$, $p_{L_1}(t)$, $p_{L_2}(t)$, $p_{L_\mu}(t)$ – values of the instant power in the equivalent circuit elements, which are determined in accordance with the [4]:

– for the resistance $p_{R_1}(t) = i^2 R_1$;

– for the inductance $p_{L_1}(t) = L \frac{di(t)}{dt} i(t)$;

– for the capacity $p_{C_1}(t) = i(t) \left(\frac{1}{C_0} \int_0^T i(t) dt \right)$;

– for the nonlinear resistance

$$p_{R_2}(t) = i^2(t) R_2$$

– for the nonlinear capacitance $p_{C_2}(t) = i(t) \left(\int_0^T \frac{i(t)}{C_2} dt \right)$;

– for the nonlinear inductance

$$p_{L_2}(t) = i^2(t) \frac{dL(t)}{dt} + L(t) i(t) \frac{di(t)}{dt}.$$

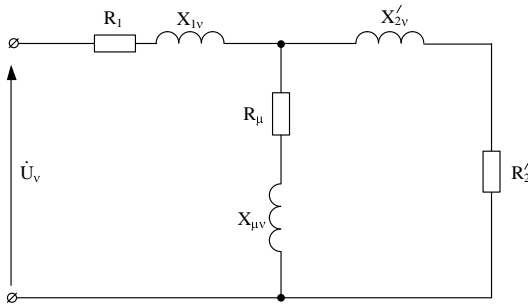


Fig. 1. Equivalent circuit of induction motor.

To solve the identification problems the voltage and the currents in the circuits are presented in the form of harmonic series. For this purpose the Fourier series, which allows us to highlight cosine and sine components of the instant power of harmonics were used.

Sinusoidal function of voltage and current in the form of truncated harmonic series could be described as [3]:

$$u(t) = \sum_{n=0}^N U_n \cos n\Omega t - \varphi_n; \quad (3)$$

$$i(t) = \sum_{m=0}^M I_m \cos m\Omega t - \varphi_m, \quad (4)$$

where: U_n – amplitude values of the constituent voltage harmonics; n – number of the harmonic voltage; N – number of voltage harmonics; Ω – the angular frequency; φ_n – phase angle between voltage and the real axis «a»; I_m – amplitude values of the constituent current harmonics; m – number of the harmonic current; M – number of current harmonics; φ_m – phase angle between voltage and current.

Expressions (3)-(4) by simple transformations could be represented as:

$$u(t) = \sum_{n=0}^N U_{na} \cos n\Omega t + \sum_{n=0}^N U_{nb} \sin n\Omega t; \quad (5)$$

$$i(t) = \sum_{m=0}^M I_{ma} \cos m\Omega t + \sum_{m=0}^M I_{mb} \sin m\Omega t, \quad (6)$$

$$\text{where: } U_{na} = U_n \cos \varphi_n; I_{ma} = I_m \cos \varphi_m; \\ U_{nb} = U_n \sin \varphi_n; I_{mb} = I_m \sin \varphi_m,$$

U_{na} , I_{ma} – cosine harmonic components of voltage and current, respectively; U_{nb} , I_{mb} – sine harmonic components of voltage and current, respectively.

Expression for the instantaneous power of the (5) and (6) [5, 6]:

$$p(t) = u(t) i(t) = \\ = \frac{1}{2} \sum_{m=1}^M \sum_{n=1}^N U_{na} I_{ma} + U_{nb} I_{mb} \cos m-n \Omega t + \\ + \frac{1}{2} \sum_{m=1}^M \sum_{n=1}^N (U_{na} I_{ma} - U_{nb} I_{mb}) \cos m+n \Omega t + \\ + \frac{1}{2} \sum_{m=1}^M \sum_{n=1}^N U_{na} I_{mb} + U_{nb} I_{ma} \sin m-n \Omega t + \\ + \frac{1}{2} \sum_{m=1}^M \sum_{n=1}^N (U_{na} I_{mb} + U_{nb} I_{ma}) \sin m+n \Omega t. \quad (7)$$

According to (7), the instantaneous power consists of the two components groups. The first group is formed by multiplying the single-frequency components of voltage and current, that is, when $m=n$. The frequencies of the instantaneous power harmonics are equal to $k_c = m-n=0$ and $k_c = m+n=2m$. The zero frequency $k=0$ corresponds to a constant value of power.

The components of the instantaneous power by multiplying single-frequency components of voltage and current are the canonical components (P_c). The components of the instantaneous power in the multiplication of different frequency components are non-canonical elements (P_s). If $m=n$ obtained the power harmonic of the canonical order k_c , and if $m \neq n$ – non-canonical order k_s .

The total value of the instantaneous power is defined as the sum of canonical and non-canonical components:

$$P_k = P_{k_c} + P_{k_s}. \quad (8)$$

With separation of alternating components power on the canonical and non-canonical expression for power can be written as [2, 3]:

$$p(t) = \sum_{k=0}^{M+N} P_{k_0} + \sum_{k=1}^{M+N} P_{ka}(t) + \sum_{k=1}^{M+N} P_{kb}(t) = \sum_{k=0}^{M+N} P_{k_0} + \\ + \sum_{k=1}^{M+N} P_{ka_c}(t) + P_{ka_s}(t) + \sum_{k=1}^{M+N} P_{kb_c}(t) + P_{kb_s}(t). \quad (9)$$

Harmonics of power (sine and cosine) with frequencies k :

$$P_{ka}(t) = P_{ka_c} + P_{ka_s} \cos k\Omega t; \\ P_{kb}(t) = P_{kb_c} + P_{kb_s} \sin k\Omega t, \quad (10)$$

where:

$$P_{ka_c} + P_{ka_s} = \frac{1}{2} \sum_{m=1}^M \sum_{n=1}^N U_{na} I_{ma} + U_{nb} I_{mb} + \\ + \frac{1}{2} \sum_{m=1}^M \sum_{n=1}^N U_{na} I_{ma} - U_{nb} I_{mb}; \\ P_{kb_c} + P_{kb_s} = \frac{1}{2} \sum_{m=1}^M \sum_{n=1}^N U_{na} I_{mb} + U_{nb} I_{ma}.$$

The values of cosine and sine components of the instantaneous power:

$$\begin{aligned} P_{ka_c} + P_{ka_s} &= \frac{2}{T} \int_0^T p \cos k\Omega t \, dt; \\ P_{kb_c} + P_{kb_s} &= \frac{2}{T} \int_0^T p \sin k\Omega t \, dt. \end{aligned} \quad (11)$$

Instantaneous power includes constant component, cosine and sine, canonical and non-canonical components. The frequencies of the instantaneous power of harmonics (k) are formed by multiplying all voltage harmonics (n) and current (m). They are defined as $k = m \pm n$. For example, the 4-th harmonic of the instantaneous power ($k = 4$) in the presence of signals of voltage and current harmonics with the numbers $m = n = 1, 3, 5, 7$ are generated by all considered harmonics, that is $k = m + n = 1 + 3 = 4$; $k = m - n = 5 - 1 = 4$; $k = m - n = 7 - 3 = 4$.

The system of equations for the T-equivalent circuit (fig. 1) is made in the form of equity components of the instantaneous power source voltage polyharmonic sum of the components of instantaneous power on the resistance and inductance [2]:

$$\begin{cases} P_{0s\Sigma} = P_{0R_l} + P_{0R_\mu} + P_{0R'_2}; \\ P_{kas\Sigma} = P_{kaR_l} + P_{kaR_\mu} + P_{kaR'_2} + P_{kaL_l} + P_{kaL_\mu} + P_{kaL'_2}; \\ P_{kbs\Sigma} = P_{kbR_l} + P_{kbR_\mu} + P_{kbR'_2} + P_{kbL_l} + P_{kbL_\mu} + P_{kbL'_2}, \end{cases} \quad (12)$$

where: $P_{0s\Sigma}, P_{0R_l}, P_{0R_\mu}, P_{0R'_2}$ – constant components of the instantaneous power of the source voltage and polyharmonic resistances of the stator, the circuit of the magnetization and rotor, respectively; $P_{kas\Sigma}, P_{kbs\Sigma}$ – cosine and sine components of the instantaneous power of the source voltage polyharmonic; $P_{kaR_l}, P_{kbR_l}, P_{kaR_\mu}, P_{kbR_\mu}, P_{kaR'_2}, P_{kbR'_2}$ – cosine and sine components of the instantaneous power in the resistance of the stator, the circuit of the magnetization and rotor; $P_{kaL_l}, P_{kbL_l}, P_{kaL_\mu}, P_{kbL_\mu}, P_{kaL'_2}, P_{kbL'_2}$ – cosine and sine components of the instantaneous power in the inductance of the stator, the circuit of the magnetization and rotor.

The equations of power balance using three voltage and current harmonics ($m = n = 1, 3, 5$):

$$\begin{aligned} P_{0s\Sigma} &= P_{0R_l} + P_{0R_\mu} + P_{0R'_2}; \\ P_{2as\Sigma} &= P_{2aR_l} + P_{2aR_\mu} + P_{2aR'_2} + P_{2aL_l} + P_{2aL_\mu} + P_{2aL'_2}; \\ P_{2bs\Sigma} &= P_{2bR_l} + P_{2bR_\mu} + P_{2bR'_2} + P_{2bL_l} + P_{2bL_\mu} + P_{2bL'_2}; \\ P_{4as\Sigma} &= P_{4aR_l} + P_{4aR_\mu} + P_{4aR'_2} + P_{4aL_l} + P_{4aL_\mu} + P_{4aL'_2}; \\ P_{4bs\Sigma} &= P_{4bR_l} + P_{4bR_\mu} + P_{4bR'_2} + P_{4bL_l} + P_{4bL_\mu} + P_{4bL'_2}; \\ P_{6as\Sigma} &= P_{6aR_l} + P_{6aR_\mu} + P_{6aR'_2} + P_{6aL_l} + P_{6aL_\mu} + P_{6aL'_2}; \\ P_{6bs\Sigma} &= P_{6bR_l} + P_{6bR_\mu} + P_{6bR'_2} + P_{6bL_l} + P_{6bL_\mu} + P_{6bL'_2}; \\ P_{8as\Sigma} &= P_{8aR_l} + P_{8aR_\mu} + P_{8aR'_2} + P_{8aL_l} + P_{8aL_\mu} + P_{8aL'_2}; \\ P_{8bs\Sigma} &= P_{8bR_l} + P_{8bR_\mu} + P_{8bR'_2} + P_{8bL_l} + P_{8bL_\mu} + P_{8bL'_2}; \\ P_{10as\Sigma} &= P_{10aR_l} + P_{10aR_\mu} + P_{10aR'_2} + P_{10aL_l} + P_{10aL_\mu} + P_{10aL'_2}; \\ P_{10bs\Sigma} &= P_{10bR_l} + P_{10bR_\mu} + P_{10bR'_2} + P_{10bL_l} + P_{10bL_\mu} + P_{10bL'_2}. \end{aligned}$$

Analysis of harmonic balance equations for the instant power shows that to determine the electromagnetic parameters of IM is sufficient to use the three harmonics of voltage and current. This give 11 components of the instantaneous power with 11 unknowns of the equivalent circuit (stator resistance R_l is determined in advance). As a source of polyharmonic voltage thyristor voltage regulator is used at angle in the range of $5\pi/12 \dots 5\pi/6$.

For evaluate the effectiveness of the energy method is used the IM ($P = 4 \text{ kW}$, $U = 380 \text{ V}$, $I_1 = 8,7 \text{ A}$, $n = 1420 \text{ rpm}$). Identification of electromagnetic parameters of the IM is lead at a fixed rotor, that is $s_r = 1$. The results of the identification of electromagnetic parameters are given in table 1.

Analysis of the results (table 1) are shows that the smallest error in the determination of electromagnetic parameters of the IM when using the energy method and does not exceed 5%. Adequacy of methods for the determination of electromagnetic parameters of the IM was determined by comparing the experimental and calculated curve of the stator currents. The degree of overlap was estimated by the coefficient of determination R^2 .

TABLE 1
RESULTS OF THE DETERMINATION OF ELECTROMAGNETIC PARAMETERS OF THE IM

Method	The electromagnetic parameters of the induction motor				R^2
	R'_2 , Ом	L_1 , Гн	L_μ , Гн	L'_2 , Гн	
Catalog parameters	1,39	0,0068	0,25	0,0067	
Error in the determination of electromagnetic parameters, %					
regulated by GOST 7217-87	1,94	36,866	29,7	7,164	0,777
energy method	1,44	2,941	2,0	1,493	0,998

IV. CONCLUSIONS

The use of components of the instant power balance equations allows to determine the electromagnetic parameters of the induction motors, which were to be repaired or have been in operation for a long time. Thus, in contrast to existing methods, the error in determining electromagnetic parameters does not exceed 5%.

Increasing the accuracy of the electromagnetic parameters of the induction motors is achieved by forming a system of equations separately for identification constant, canonical and non-canonical components of instant power.

REFERENCES

- [1] G.G. Rogozin, The electromagnetic parameter of induction motors determination. New experimental methods. K.: Tehnika, 1992, 168 p. [in Russian]
- [2] D.Y. Rodkin, A.P. Kalinov, Y.V. Romashihin, Energy method to identifications parameters of induction motors, Bulletin of the KDPU, Kremenchuk: KDPU, 2007, Iss. 3(44), PP. 130–136. [in Russian]
- [3] D.Y. Rodkin, The terminological notions when use the device to instant power, Lugansk Works branches to International Academy of information, Lugansk, 2005, №1(10), PP. 145–154. [in Russian]
- [4] D.Y. Rodkin, Y.V. Romashihin, Instant power of nonlinear circuit elements, Bulletin of the DGTU, Dneprodzerzhinsk: DGTU, 2007, PP. 501–506. [in Russian]
- [5] H. Akagi, E.H. Watanabe, M. Aredes, Instantaneous power theory and applications to power conditioning, Published by John Wiley & Sons, Inc., Hoboken, New Jersey, 2007. – 379 p.
- [6] A.F. Krogeris, K.K. Rashevits, E.P. Treymanis, Y.K. Shinka, AC power. Riga: Physical-energycal. Institute of Latvia Academy of Sciences, 1993, 294 p. [in Russian]

Transient stability of generator in simple power system

¹Pavol HOCKO (2nd year), ²Matúš NOVÁK (1st year)
Supervisor: ³Michal KOLCUN

^{1,2,3}Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic

¹pavol.hocko@tuke.sk, ²matus.novak@tuke.sk, ³michal.kolcun@tuke.sk

Abstract—This paper discuss about transient stability in a Power System. A modeled grid represents equivalent of the power system with infinity power and separated generator connected to this grid. Modeled generator (G1) is connected to this ideal power system trough the double transmission line. Generator's transient stability parameters are monitored during three phase short-circuit on transmission line. The goal is to find a fault critical clearing time (CCT) using surface method. The second part of the paper consists of power angle comparison in different fault times.

Keywords—CCT, exciter, fault, generator, MODES, power angle, power system, short circuit.

I. INTRODUCTION

Consideration of Steady-state stability should be used only in small phase oscillation i.e. in case of small power angle changes. Many rapid changes in the power system are caused by an imbalance between generation and consumption of real power. These changes can cause a large power angle changes. The significant changes are switching operation (e.g. transmission line or transformer switching), sudden load and especially short circuits. All these cases cause a step change in impedance. Because of the inertia of a turbine-generator set a generator's performance values (e.g. power angle – delta) are during the impedance changes constant. This phenomenon causes electromechanical oscillations. The oscillation process option can be small (power angle steady in time) or large (power angle continuously increases). Second option means generator instability, loss of synchronization and finally generator disconnection. The turbine driving torque cannot be step changed. The reason for this is mechanical control system of turbine-generator set inertia.

According to (1) step change in coupling impedance causes step change in generator output power.

$$P_1 = 3E^2 Y_{11} \sin \alpha_{11} + 3EUY_{12} \sin(\vartheta - \alpha_{12}) \quad (1)$$

where: E – alternator inductive electromotive voltage [V], Y_{11} – element of node admittance matrix, element of alternator coupling matrix [1/Ω], α – voltage phasor angle [degrees], U – voltage phasor [V], θ – alternator power angle (angle between the phasor of electromotive voltage and the phasor of generator voltage [degrees])

Because the mechanical controller lack the mechanical power generator input is consider as constant during the short transient electromagnetic disturbance. This fact is valid only in the case of short disturbances (less than 1 second).

Exciter controller influence can be neglected in initial transient state. Because of this fact, we use exciter controller settings with constant exciting (i.e. CONS exciter type in “MODES”).

II. DEFINITION OF DYNAMIC STABILITY METHOD

A. Area rule

The power system transient stability determination is a significant task which can define the electro mechanical transient process. This process is result of each significant change in the power system.

The area rule is one of the most widely used methods. This method is suitable to simply and practical determination of a critical power angle. Exact critical clearing time (CCT) determination (after a short circuit) is most important in the field of power system operation. To determine this time is important to know the load angle timing. To determine this motion differential equation of synchronous machine is widely used [8].

$$\frac{d^2 \vartheta}{dt^2} = \frac{\omega_0 \Delta P}{T_m S_n} \quad (2)$$

where: θ – alternator power angle [degrees], t – time [sec], ω_0 – synchronous rotational speed [rad/sec], ΔP – power difference (accelerating or decelerating) [MW], T_m – mechanic time constant [sec], S_n – generator nominal apparent power [MVA]

The most significant transient changes in the power system (in the area of binding impedance) are caused by short circuits. The binding impedance drop depends on the place and type of the short circuit. Three-phase near short circuit results in a drop to zero of transmitted power. The synchronous generator unloading due to failure in the power system is a serious problem which can cause dynamic stability loss in the power system.

B. Generator unit

Figure 1 shows the simplified scheme of power generation in generator unit. The synchronous generator is powered by the prime mover which is usually a turbine or a diesel engine. The primary mechanical unit is equipped with controller which regulates speed or generating power, depended on predetermined mode. A steam boiler or a nuclear reactor is source of steam for a steam turbine. The grid is powered from the generator unit through a block transformer. Additional essential elements in generator unit are exciter, exciter controller and measuring equipment. Another important subsystem is self-consumption which is powered from self-consumption transformer.

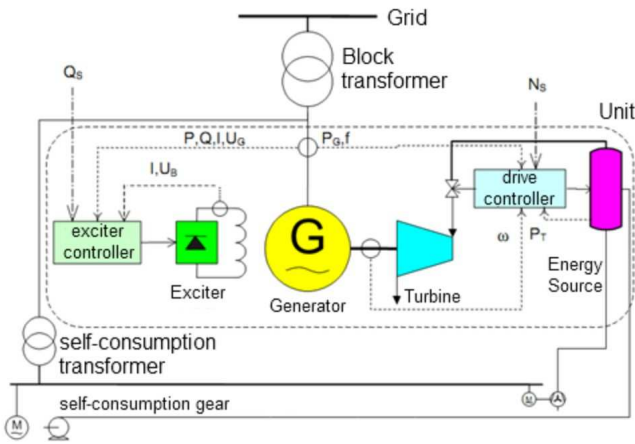


Fig. 1. Simplified generator unit scheme with self-consumption

For the purpose of creating a dynamic model in “MODES” term “generator block” will be used for whole generator unit include its prime mover, energy source and both controllers.

C. Generator models

Synchronous generators can be divided into a high speed – turbo generators driven by steam-turbine or gas-turbine and a low speed driven by water-turbines. All types of generators have a stationary part which creates a rotating magnetic field (stator) and a rotary part with field winding (rotor). The field winding is powered by the exciter.

The magnetic field is created by DC excitation current. Generated magnetic flux induces electromotive force which creates current and responsive electric power in each of three phases.

Adjusted Park's equations are used to calculate electromagnetic response in “d-q” coordinate system. Equations (3) to (6) according to [5] and [6] are given to “d” and “q” axis.

$$T_{d0}' \times E_q' = U_b + (X_d - X_d') \times I_d - E_q' \quad (3)$$

$$T_{q0}' \times E_q' = -(X_q - X_q') \times I_q - E_d' \quad (4)$$

$$T_{d0}'' \times E_q'' = E_q' + (X_d' - X_d'') \times I_d - E_q'' \quad (5)$$

$$T_{q0}'' \times E_d'' = E_q' - (X_q' - X_q'') \times I_q - E_d'' \quad (6)$$

where: I_d , I_q , and U_b – stator currents and exciting voltage [A, V], T_{d0}' , T_{d0}'' , T_{q0}' – time constants in no load condition [sec],

X_d , X_d' , X_d'' and X_q , X_q' , X_q'' – synchronous, transient and impact reactance in direct and cross-wind axis [Ω]

If transient response in the rotor excitation circuits is neglected and electromotive forces (E_q' and E_d') are considered as constant a “classic model” of the generator is created. This generator model is defined as “CLAS” in program “MODES” and it is suitable for a middle-time dynamic and generators which are electrically distant from faults modeling.

If a slip (s_G) and the power angle (δ) are also considered as constant a “constant model” of generator is created. This generator model is defined as “CONS” in program “MODES” and it is suitable for transient state modeling in the power system which is interconnected to another power system with infinity power. This model is used to modeling a power system connected to modeled generator G1. [2]

D. Exciter models

The generator's exciter consists of exciter itself and the exciter controller. An exciter power is generally 0.2 to 0.8 % of generator power and an exciter voltage is below 1 kV due to additional winding insulation. Exciters can be divided to rotary and stationary. Rotary exciters can be divided to DC exciters and AC exciters equipped with a rectifier. [2] Figure 2 shows the static autonomous exciter. This type of exciter is defined as “AC_4” in program “MODES”.

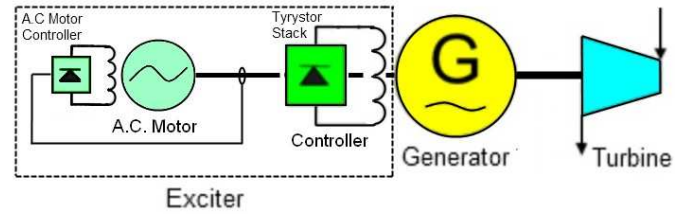


Fig. 2. Static autonomous AC exciter, marked as “AC_4”

III. THREE-PHASE SHORT CIRCUIT MODELING

A. Modeled grid

Figure 3 shows a single-pole four node grid according to [3] which is modeled in program “MODES”. Generator (G1) generates rated power 588 MW, rated voltage 20 kV and rated power factor $\cos\phi$ 0.85. The generator G1 is connected to the power system with infinite power (INFBUS) through a block transformer (T1), a 200 km long double transmission line and a distribution transformer (T2). The INFBUS is modeled as the “CONS” generator block type connected to node 4.

The project created in “MODES” monitors a generator's dynamic stability in case of the three-phase short circuit which occurs on V2 transmission line. Each case represents different fault time scenario.

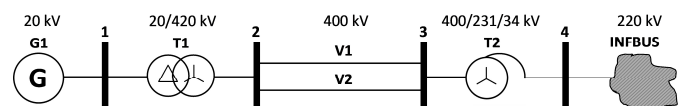


Fig. 3 Modeled four node grid with monitored generator G1

The area rule and dynamic simulation are used to find the critical clearing time (CCT). The CCT is maximum duration of short circuit when is the generator still stabilized and synchronized with the power system after fault.

Step one rests in modeled grid parameters importing in to “MODES”. The grid consists of nodes, branches and generators. These elements have to be set individually in a operation editor section. Second step is generator’s dynamic model definition. As has been said in Section B – Generator unit, each generator consists of energy source, turbine, generator itself, exciter, exciter controller and other controllers. Each of these elements has own parameters which have to be defined. The generator dynamic parameters are set in dynamic operation section.

The final step is output graphic definition. Dynamic parameters of generator will be monitored. Following parameters will be set. power angle delta (angle between rotor “q” axis and synchronously rotating voltage phasor axis) – “DELT”, generator real power – “PG”, accelerating torque – “AK” and generator slip (proportional deviation between rotor speed and synchronous speed) – “SG”

B. Case 1- fault time set to 0.1 sec

In the case 1 the short circuit time duration is set to 0.1 sec. Short circuit initialization time is set to 0.1 sec. and fault disconnection is set to 0.2 sec. The results are in figure 4.

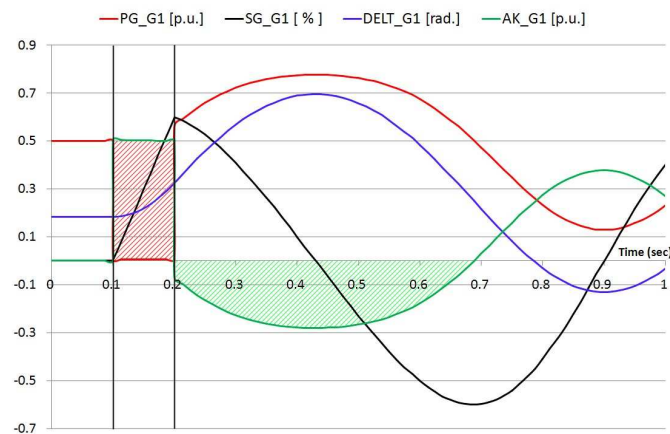


Fig. 4 Monitored dynamic values of generator in fault with 0.1 s time duration

The generator dynamic process is as follows. The generator’s real power (PG, in chart marks as red line) drops to zero due to short circuit in time of 0.1 sec. The turbine mechanical power is constant. The generator starts accelerate as consequence of positive acceleration torque (AK, in chart marks as green line). The generator’s slip (SG, in chart marks as black line) has approximately linear rising. The power angle delta (DELT, marks as blue line) rises quadratically. Fault shutdown in time of 0.2 sec. is in chart marked as vertical line. The generator’s electric power steeply changes value and gets over original value before the fault. Generator is decelerating. By the reason that accelerate area (highlight as red surface) is smaller than decelerate area (highlighted as green surface) electromechanical oscillation is stabile and generator stays synchronously connected to the grid.

C. Case 2- fault time set to 0.2 sec

In the case 2 short circuit time duration is set to 0.2 sec. Short circuit initialization time stays set to 0.1 sec. and fault disconnection is now set to 0.3 sec. The results of the real electric power, the acceleration torque power angle and the slip are in figure 5. The red acceleration area is bigger than the green deceleration area in this case. Generator rotor is accelerating as far as it gets off from synchronism. The electromechanical oscillation is unstable and the machine goes to an asynchronous state. Generator real power curve has typical oscillative shape.

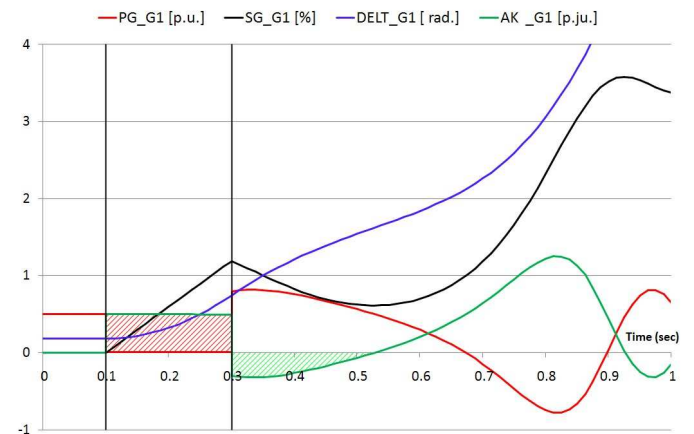


Fig. 5 Monitored dynamic values of generator in fault with 0.2 s time duration

This unstable asynchronous state is very dangerous for the power system dynamic transient stability. The power oscillation after the fault mechanically stress generator and can damage or even destroy the generator. In the event of this case generator has to be rapidly disconnected from the grid.

D. Case 3 – without fault shutdown

The third case deals with scenario where affected transmission line is not disconnected. The generator real power is zero after the short circuit initialization. Generator slip rises linearly and power angle rises exponentially. This is unstable state. The power angle curve compared with other cases is in figure 6.

E. Case 4 – CCT

In the last case critical clearing time is searched. Due to precious cases CCT will be between 0.1 and 0.2 sec. time duration. In the CCT finding scenario we need to regard that time in simulation is not continuous function. Time simulation in “MODES” has particular time steps. The smallest time stamp is equal to an integral constant. We have to set the integral constant to $t_i = 0.00625$ sec. to find CCT. The CCT finding process is very simple. We will increase short circuit time duration to time when the generator G1 gets off from the synchronism (accelerating area will be bigger than decelerating). In our case CCT has value of 0.18125 sec. Figure 6 shows generator values in the case when the transmission line is disconnected at the CCT. The Accelerating area is slightly smaller than the decelerating area.

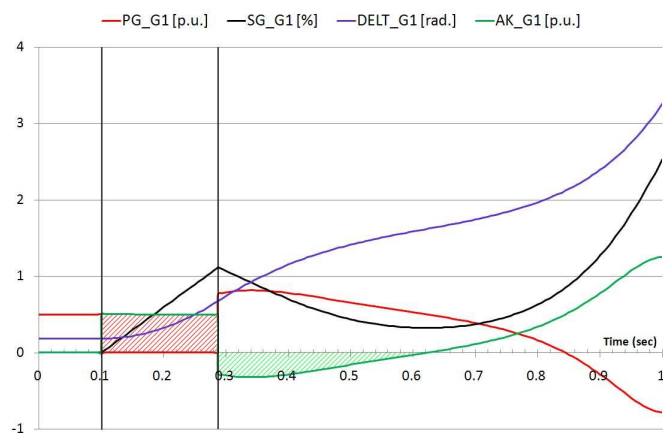
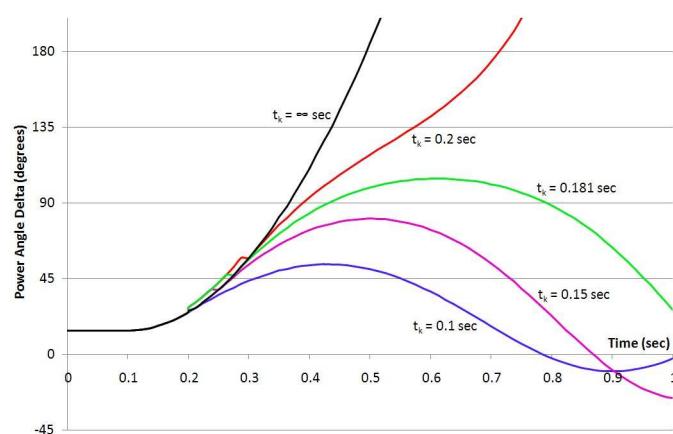


Fig. 6 Monitored dynamic values of generator fault in CCT

According to the area rule generator operation is stable after the fault. Figure 7 shows the power angle comparison in different fault times.

Fig. 7 Power angle - delta in different fault time t_k

As figure 7 shows fault time duration has effect to power angle. Black curve shows the power angle in the Case 3 when fault was not disconnected. Green curve shows the power angle in case when fault was disconnected at the CCT. The green curve is decreasing after time of 0.6 sec. which means stable state of the generator as was said before.

IV. CONCLUSION

An influence of short circuit was shown on the four node grid. This data was obtained from the dynamic simulation in program “MODES”. In order to simplification synchronous machine was modeled as a motor with a constant electromotive force (constant inductive voltage). The exciter was set as constant. The maximum stable fault time – CCT was found in 0.18125 sec. The last part of project was the power angle comparison. This comparison visually demonstrated relation between the fault time duration and the power angle. The CCT may also depend on the generator's and the exciter's type parameters. In the first place exciter amplification constant which is product of gain constant and proportional gain constant is important. Decreased resultant exciter amplification has direct impact to CCT.

ACKNOWLEDGMENT

This publication is the result of the Project implementation: Research centre for efficient integration of the renewable energy sources, ITMS: 26220220064 supported by the Research & Development Operational Programme funded by



We support research activities in Slovakia / Project is cofinanced from EU funds.

REFERENCES

- [1] Z. Trojáněk, J. Hájek, P. Kvasnica “Přechodové jevy v elektrizačních soustavách” 1987 pg. 202–212
- [2] K. Máslo „Tvorba dynamických modelů pro praktické výpočty, seminár Aktuální otázky a vybrané problémy řízení ES“. Dostupné na internete www.modesinfo.com
- [3] V. Chladný, M. Bilička, “ Přechodové jevy v elektrizačních sítích” November 1991
- [4] J. E. Gurevi “Rasčoty ustojčivosti a protiviavarnijnoj automatiky a v energosistemach” Energoatomizdat 1990
- [5] J. Arrillaga & col.: Computer Modelling of Electrical Power System; John Wiley & Sons ; 1983
- [6] J. Machowski, J.W.Bialek, J.R.Bumby: Power System Dynamics and stability, John Wiley & Sons ; 1997
- [7] H. Saadat „Power System Analysis“ 2004
- [8] M. Kolcun, V. Griger, L. Bena, J. Rusnak: „Analýza elektrizačnej sústavy“ Košice 2005
- [9] K. Máslo: Influence of wind farms on transmission system operation in the central Europe, the 9th International Conference Control of Power Systems, Tatranské Matliare, Slovakia, 2010., ISBN 978-80-89409-19-9
- [10] Kušník, S., Beňa, L., Kolcun, M.: The Impact of FACTS Devices to Control the Load Flow. In Proceedings of the 11th International Scientific Conference Electric Power Engineering 2010. Brno: Brno University of Technology, 2010, vol. 1, p. 99-103. ISBN 978-80-214-4094-4.

Transient stability of power system

¹Matúš NOVÁK (1st year), ²Pavol HOCKO (2nd year)
Supervisor: ³Michal Kolcun

^{1,2,3}Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

¹matus.novak@tuke.sk, ²pavol.hocko@tuke.sk, ³michal.kolcun@tuke.sk

Abstract— Today's large scale power systems became more complicated and complex. Also human's dependence on electricity increases rapidly in last years. Power system outages have large economical and social impact. One of the main goals which power system operators have to accomplish is to minimize the effect of disturbances in the power system so as to have a minimal impact on reliability and security the of power supply. Power system stability is one of the important issues which have to be studied more precisely to reach this goal. The loss of stability can lead to outages of transmission lines, loss of loads, loss of generation, cascading outages and eventually to black-out. This paper focuses mainly on the basics of the power system stability, and the need of stability analysis.

Keywords—power system, renewable energy sources, steady-state stability, transient stability, PSLE.

I. INTRODUCTION

Power system as a large dynamic system of cybernetic type includes many interconnected components, operated at a large area with significant feedbacks. Considering that fact, one of its main characteristics is complexity of mutual influences of components. With interconnecting of the power systems of particular states into international interconnected power system, this complexity factor also grow. On the other side, interconnected power systems have many advantages, for example they are more resistant to some types of disturbances. With opening of the power market and with larger number of renewable energy sources (RES) connected to the power system, some new advantages, but also new threats, appeared.

For this reason, it is important to analyze the transient stability of the power system, because the overloading of transmission lines and other equipment of the power system appears more often due to growing rate of electrical power produced by RES.

II. BASIS FOR STABILITY CALCULATIONS

During the power system operation, two states can occur, first, defined as steady-state, and second, transient state. Under steady-state conditions we understand the state of equilibrium of electrical and mechanical variables of power system. Every change of these variables causes the appearance of transient state. In interconnected network, rotors of all generators run synchronously with the same synchronous electrical speed. At the same time, mechanical power (torque) input supplied by the prime mover and electrical power (torque) output (neglecting losses) are in balance, as shown on Fig. 1. To meet

these conditions, all values of power system operating parameters as frequency, voltage, real and reactive power, at all power system equipment, must comply with steady-state balance condition.

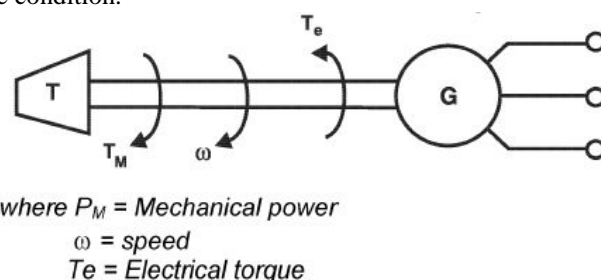


Fig. 1. Power balance condition in power generator [4].

Power system in steady-state can be described by the system of nonlinear algebraic equations. Because of power system complexity, it is convenient to use one of numeric methods for solving this system of equations, as Newton, Newton-Raphson or Gauss-Seidel method, implemented on digital computers. Values of operating parameters in steady-state are an initial condition for analyze power system stability.

Transient state is characterized as a state when power system due to imbalance one of operating parameters, change its state from one steady-state to another, or when this imbalance is so large, that system stability couldn't be maintained. System disturbances in terms of duration can be classified into three main categories:

- Wave disturbances with approximate duration from micro- to milliseconds
- Electromagnetic disturbances with approximate duration from milli- to seconds
- Electromechanical disturbances with approximate duration from 0,1s to minutes.

Disturbances significantly affect power system stability, therefore it is important to be familiar with behavior of system during disturbances to prevent operating power system in critical conditions.

Power system stability is then defined as ability of power system to restore steady-state conditions to state before disturbance, or take new steady-state, after one or more changes of operating parameters. We can classify power system stability into number of categories; best known classification distinguishes steady-state and dynamic stability. IEEE and Cigre classify power system stability according to observed operating parameter as shown on Fig. 3.

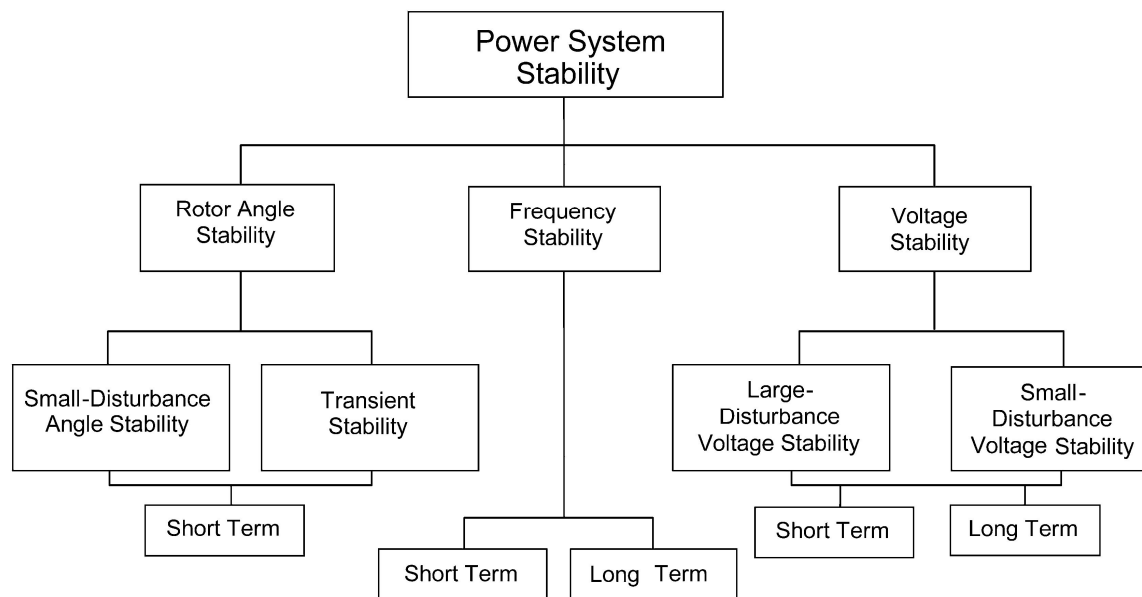
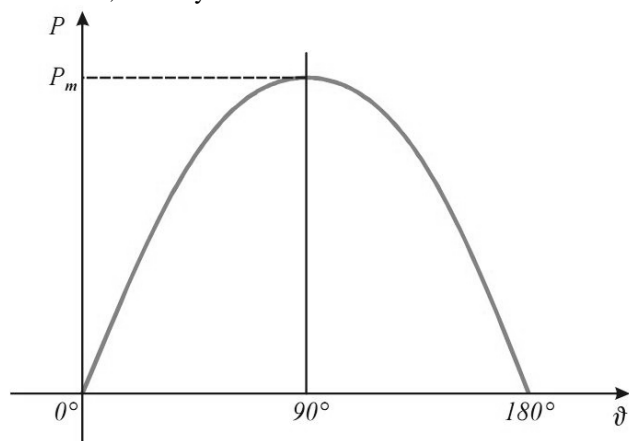


Fig. 3. Classification of power system stability according IEEE [21].

III. STEADY-STATE STABILITY

By the term of steady-state stability we understand ability of power system to stay in the same or very similar steady-state operational conditions after any small disturbance. Steady-state stability involves phenomena whose common characteristic is a small change in operating parameters which may lead to stability or instability of power system. There are two options after this small change of parameters. If power angle δ during disturbance does not exceed 90° , stability will be maintained and power generator angle will get back to its value before disturbance. However, when power angle exceeds 90° , stability will not be maintained.

Fig. 2. Dependency of output electrical power P on power angle δ [1].

The most common deviations regarding the generator from normal operating conditions, which threaten steady-state stability are:

- Decrease in field voltage in generator (incorrect operation of exciter regulator)
- Sudden decrease (increase) in load in power system load bus (slow load change in power system)
- Sudden increase (decrease) in turbine mechanical power output (incorrect operation of turbine governor)

Common sign for all these disturbances is that the unbalance of mechanical power supplied by the turbine (containing also losses in generator) and electrical output

power occurs, resulting in appearance of accelerating (of decelerating) power which causes rotor acceleration (or deceleration). This power is a threat for the power system stability, because generator may not be able to maintain stability.

Solving steady-state stability of whole power system includes large number of generators, transformers, transmission lines and other devices. Due to complexity of this task, digital computers and computer software are used. More frequent task than solving whole power system is solving steady-state stability of small particular system connected to power system. There are also tasks which aim to check whether synchronous operation of particular generation plant is possible or not. Different simplifications are used, mostly used is reduction of a solved part of power system to one generator and *infinite bus* which is defined as bus with constant frequency and voltage.

IV. TRANSIENT STABILITY

Generators are interconnected by network consisting of lines, substations and other devices. Behavior of this network during disturbances can be described by an interesting mechanical analogy [5]. As shown on Fig. 4, a number of masses in this analogy, for example steel balls, representing synchronous generators are interconnected by network of rubber strings, representing transmission lines. We assume that the network is in steady-state, and the force on each string is below its break point. Cutting one of rubber strings represents loss of one of the transmission lines. As a result, masses undergo transient oscillations and the forces on the strings fluctuate. There are two possibilities. In the first case, the system will settle down to a new steady-state with a new set of string forces. In the second case, transient oscillations will cause disruption of other strings, resulting in even weaker network, which is more vulnerable to oscillations and eventually system collapse.[5]

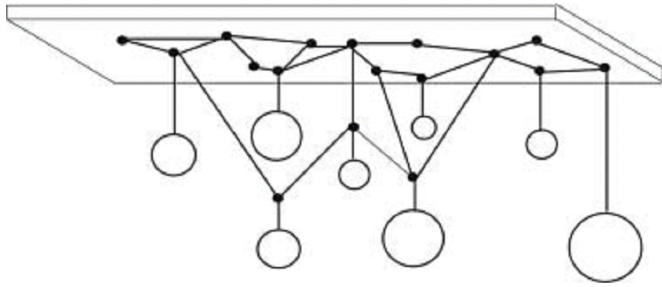


Fig. 4. Mechanical analogy of transient stability [5].

As for the theory of power systems, the transient stability phenomena include all sudden changes of large amplitude. These changes are: unbalance in mechanical input and electrical output power of generator, transient deviations from power frequency and generator power angle changes. These sudden changes are caused by major disturbances such as switching operations, sudden load changes, the loss of generation, faults, especially short circuits.

Generally speaking, these phenomena include step changes of load impedance. Power systems undergo transient oscillations, because due to inertia of rotating masses in generators, operating parameters could not change immediately, especially the power angle δ . There may be two different results of these oscillations. In the first case, the power angle will settle down at a new constant value. In the second, worse case, the power angle will value continue to increase, which may result in instability of power system or only specific generator, when we are talking about model of simple power system with one generator connected to an infinite bus. Stability or instability of the power system after disturbance depends on *critical clearing time*, for example clearing time of a short-circuit on transmission line. When clearing time exceeds critical clearing time for specific location in the power system, transient stability may not be maintained for some generators in the power system. The value of critical clearing time may be obtained by solving differential swing equation.

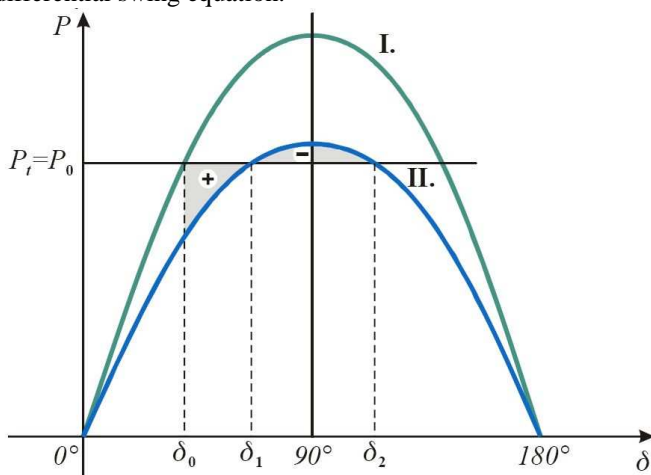


Fig. 5. Equal area criterion [1].

Another method for determining transient stability is *equal area criterion* which principle is shown on Fig. 5. This method is applicable for one machine connected to an infinite bus or for two machines. Using this method, we can obtain value of *critical power angle*, from which can obtain the value of critical clearing time. This method is based on the

comparison of two areas called *accelerating area* and *decelerating area*. Size of these areas can be obtained by integrating of modified swing equation within the limits of δ_0, δ_1 , or δ_1, δ_2 , respectively. Stability condition is then:

$$S^+ \leq S^- \Rightarrow \int_{\delta_0}^{\delta_1} \Delta P d\delta \leq \int_{\delta_1}^{\delta_2} \Delta P d\delta \quad (1)$$

Respecting this statement, we can define the critical power angle δ_k as angle to which limit condition can be applied:

$$S^+ = S^- \quad (2)$$

V. MODELING OF TRANSIENT PHENOMENA IN PSLF

To evaluate the power system state in steady-state and also in transient conditions, it is convenient to use computer software. One of several software solutions is U. S. software Positive Sequence Load Flow or simply PSLF, from General Electric. Another example of software is Modes, a Czech software.

PSLF allows modeling of all components of the power system including generators, exciters, governors, controllers, relays, loads etc. PSLF is designed for the complex analysis of the power system including calculation of short circuit currents, and analysis of the steady-state and transient stability. In this chapter, we show an example of using PSLF to model transient stability phenomena on a simple power system with eight buses and two generators.

In PSLF, power system is described by the system of algebraic equations and differential equations of generators and other devices. Differential equations are given by dynamic model library. Individual models are connected to the corresponding bus. All dynamic simulations use model library for reading and saving model parameters, and then performing calculations and sending the output values to the results file. This results file is given by the outputs of all connected dynamic models.

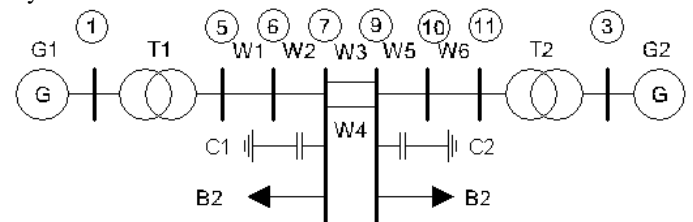


Fig. 6. Modeled power system.

Consider power system shown on Fig. 6. Parameters of power system components are the following:

generators:

$U_n = 20$ kV, $P_n = 900$ MW, $x_d = 1,8$, $x_q = 1,7$, $x'_d = 0,3$, $x'_q = 0,55$, $x''_d = 0,25$, $x''_q = 0,25$, $x_1 = 0,2$, $r_a = 0,002$, $T'_d = 8$ s, $T''_d = 0,4$ s, $T'_q = 0,03$ s, $T''_q = 0,05$ s,

power transformers:

$U_{nh} = 230$ kV, $U_n = 20$ kV $S_N = 1200$ MVA, $R = 0.0018$ p. j. , $X = 0.025$ p. j. , $B = -0.0054$ p. j.

lines:

W1, W6 : $U = 220$ kV, $R = 0.002$ p. j. , $X = 0.025$ p. j. , $B = 0.0043$ p. j.

W2, W5 : $U = 220$ kV, $R = 0.001$ p. j. , $X = 0.010$ p. j. , $B = 0.0017$ p. j.

W3, W4 : $U = 220$ kV, $R = 0.022$ p. j. , $X = 0.220$ p. j. , $B =$

0.385 p. j.

loads:

$S_{B1}=475+j111$ MVA, $S_{B2}=945+j114$ MVA, $Q_{C1}=180$ MVar, $Q_{C2}=200$ MVar.

For modeling generators is possible to use various types of dynamic models, such as models for steam-turbine generators, hydro-generators, generators for wind turbines, combined cycle generators, etc. In this case, we choose model "genrou:" for both generators. Block diagram is shown on Fig. 7.

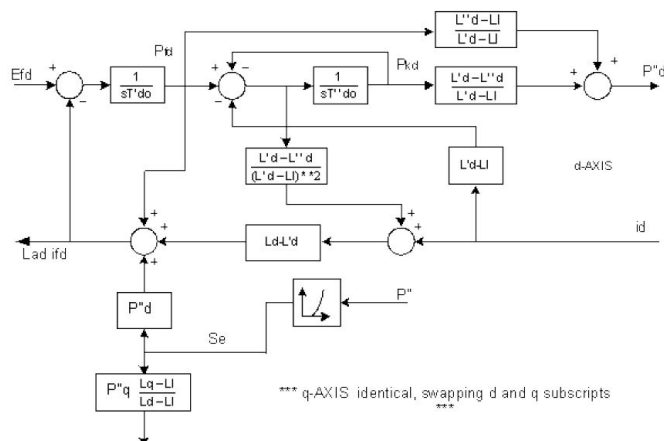


Fig. 7. Block diagram of "genrou" dynamic model in PSLE [6].

Now we consider that the fault occurs on one of the parallel lines between the bus 7 and 9, nearby bus 9, at the time 0,2 s, and the fault would be cleared after 0,5 s, at the time 0,7, by opening circuit breakers on both sides of the line. Furthermore, we don't consider any reclosing operations.

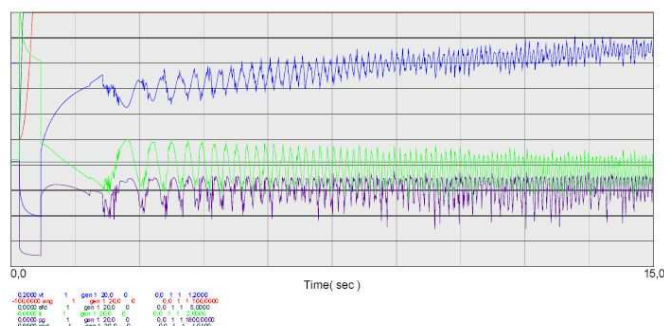


Fig. 8. PSLE output plots for generator 1

The generator no.1 parameter curves are displayed on Fig. 8 (rotor angle, terminal voltage, field winding voltage, terminal current, active and reactive output power). Fig. 8 shows that oscillations occur on the generator because machine regulator is unable to stabilize the generator.

VI. CONCLUSION

Power system stability phenomena are in current conditions of increasing consumption of electricity, connecting renewable energy sources, and other issues. It becomes matter of critical importance, mainly due to the existing lines and machines are often run up to their stability limit. It is therefore important to have a precise analysis of the power system, which is allowed by the use of a software such as PSLE.

ACKNOWLEDGMENT

This publication is the result of the Project implementation: Research centre for efficient integration of the renewable energy sources, ITMS: 26220220064 supported by the Research & Development Operational Programme funded by the ERDF.



We support research activities in Slovakia / Project is cofinanced from EU funds.

REFERENCES

- [1] Trojánek, Z – Hájek, J – Kvasinca, P: *Přechodné jevy v elektrizačních soustavách*. Praha: 1. vydání. Vydavatelstvo SNTL/ALFA 1987. 04-534-87
- [2] Kundur, P. - Paserba, J. - Ajarapu, V. - Andersson, G. - Bose, A. - Canizares, C. - Hatziaargyriou, N. - Hill, D. - Stankovic, A. - Taylor, C. - Van Cutsem, T. - Vittal, V.: *Definition and classification of power system stability*. In: Power Systems, IEEE Transactions on , roč.19, č.3, s. 1387- 1401, Aug. 2004. ISSN 0885-8950
- [3] Krištof, V. – Kušnir, S. – Hlubeň, D. – Kolcun, M.: *Modelovanie prechodných javov v PSLE*. In: Elektroenergetika. Roč.3, č.6, s.13-16. ISSN 1337-6756
- [4] Krištof, V. - Kušnir, S. - Katin, M.: *Stability of Power System*. In: SCYR 2011: 11th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice : proceedings from conference : 17th May 2011, Herľany Slovakia. Košice : FEI TU, 2011 S. 177-180. ISBN 978-80-553-0644-5
- [5] Glover, J. – Sarma, M. – Overbye, T.: *Power System Analysis and Design*. Fifth Edition, Global Engineering, USA, 2010. ISBN-13: 978-1-111-42577-7
- [6] Software documentation PSLE 17.0_05. General electric 2009.
- [7] Vittal, V: *Transient stability and control of Large Scale Power Systems*. In: PSERC Background Paper, Department of Electrical and Computer Engineering, Iowa State University. Iowa, 2003.

Using the TCSC for power flow control and reduce power losses in power system

¹Roman JAKUBČÁK (1st year), ²Matúš KATIN (3rd year)
Supervisor: ³Lubomír BEŇA

^{1,2,3}Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic

¹roman.jakubcak@tuke.sk, ²matus.katin@tuke.sk, ³lubomir.bena@tuke.sk

Abstract— Article discusses the use of FACTS (Flexible Alternating Current Transmission System) devices in power systems. These are devices that are able to provide an increase in transmission capacity and controllable power systems. The article will be provided for a particular application FACTS devices to control power flow on a simple model of the power system. Followed by monitoring the impact of regulation on the size of the active power losses.

Keywords—power system, thyristor controlled series compensator, power losses, power flow.

I. INTRODUCTION

Today we are witnessing a continuous increase in electricity demand. This trend, along with market liberalization cause problems in the management of power systems (PS).

For this reason, more and more increasing requirements on possible ways to control PS. One of these options is a Thyristor controlled series compensator (TCSC). TCSC provides the ability to manage the power flows in PS as well as an additional feature the ability to reduce power losses in PS, thus saving resources from which electric power is obtained. Thus contribute to protecting the environment and reduces the production of CO₂.

II. TCSC

TCSC configurations use thyristor controlled reactors (TCRs) in parallel with segments of a capacitor bank. The TCSC combination allows the fundamental frequency capacitive reactance to be smoothly controlled over a wide range and switched to a condition where the bi-directional thyristor pairs conduct continually and insert an inductive reactance in the line. It is alternative to static synchronous series compensator (SSSC)[1].

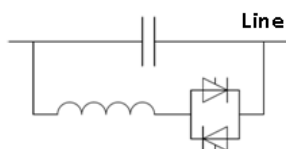


Fig. 1. Thyristor controlled series compensator.

Transmitted power is determined as:

$$P = \frac{U_1 \cdot U_2}{X_{line}} \cdot \sin(\delta_1 - \delta_2) \quad (1)$$

Where P is transmitted power, U₁ and U₂ are the voltages at the beginning and end of line, X_{line} is the line reactance, angles δ₁ and δ₂ are the voltages angles at the beginning and end of line. After addition of TCSC in line changes:

$$P = \frac{U_1 \cdot (U_2 \pm \Delta U)}{X_{line} \pm \Delta X} \cdot \sin(\delta_1 - \delta_2 \pm \Delta \delta) \quad (2)$$

Where ΔX is added reactance of TCSC and Δδ is change in voltage angle affects of TCSC.

Advantages of the TCSC[2]:

- Rapid and continuous of the transmission line series compensation level.
- Dynamic control of power flow in selected transmission lines.
- Prevent the loop flow of power
- Damping of the power swings from local and inter-area oscillations.
- Enhance level of protection for series capacitors. (A fast bypass of the series capacitor during faults can protect capacitor).
- Voltage support.
- Reduction of the short-circuit current.

Modes of TCSC operation[2]:

- Bypassed thyristor mode - in this bypassed mode, the thyristors are made to fully conduct with a conduction angle of 180°. The TCSC module behave like a parallel capacitor – inductor combination.
- Blocked thyristor mode - in this mode the firing pulses to the thyristor valves are blocked. TCSC reactance is capacitive.

- Capacitive mode - in this mode the device operates in capacitive operation.
- Inductive mode - TCSC operates in inductive operation.

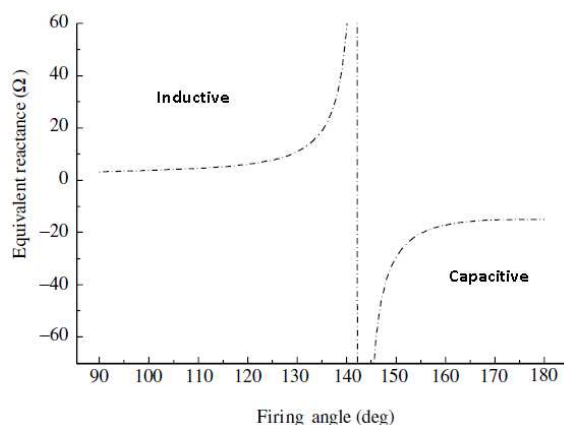


Fig. 2. TCSC fundamental frequency impedance[3].

TCSC model consists of a variable impedance connected in series to the line. In common operation we can compensate line for 40% to 60%. Then, total impedance line where is TCSC connected:

$$X_t = X_{line} + X_{TCSC} \quad (3)$$

Where X_t is total line reactance, X_{line} is line reactance and X_{TCSC} is added reactance TCSC.

III. POWER FLOW CONTROL

The following example is shown how we can change the active power flow using the TCSC. To change the opening angle thyristors TCSC impedance changes and the device operates in the inductive or capacitive mode. In the inductive mode impedance increase and the line will decrease the transmitted active power. In the capacitive mode, the inductance of the line compensated by TCSC, thus resulting impedance is lower than without TCSC, which lead to an increase in transmitted active power.

The following figures shows a simulated network, changes in power flow of individual lines and total active power loss across the network depending on the degree of line compensation, in which the TCSC is installed.

All simulation was performed in program MATLAB.

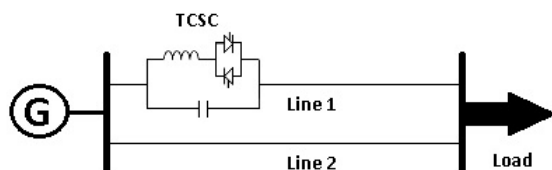


Fig. 3. Two parallel lines with TCSC.

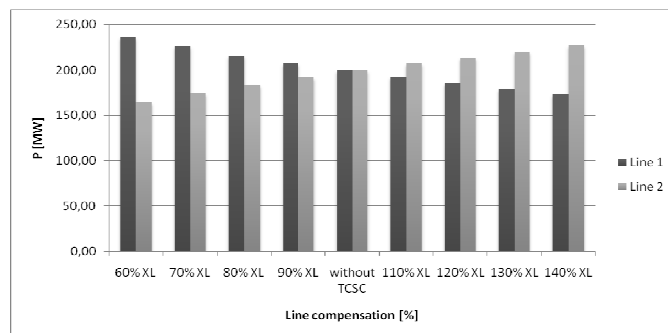


Fig. 4. Active power flow changes depending the setting of TCSC.

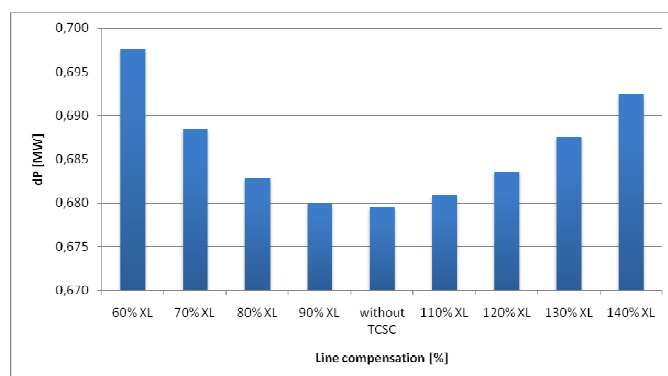


Fig. 5. Change in total power losses.

As seen from these graphs, the total active power losses increase with the degree of line compensation change and not decrease when the device operates in capacitive mode and reduces the impedance of the line in which it is located.

IV. USING TCSC IN POWER SYSTEM TO REGULATE POWER FLOW AND REDUCE LOSSES

In the considered power system in fig. 6, TCSC is connected in line 11 (from node 1 to node 10). In the program Matlab was this location evaluated as the optimal location in terms of reducing total power losses.

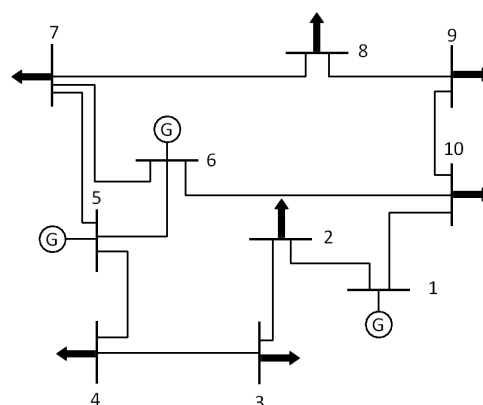


Fig. 6. Change in total power losses.

TABLE I
GENERATIONS AND LOADS IN NODES

Node	Generation		Load	
	P [MW]	Q [MVar]	P [MW]	Q [MVar]
1	612,325	514,829	0	0
5	500	0	0	0
6	1000	300	0	0
2	0	0	300	150
3	0	0	300	150
4	0	0	300	150
7	0	0	300	150
8	0	0	300	150
9	0	0	300	150
10	0	0	300	150

TABLE II
POWER FLOW AND TOTAL POWER LOSSES IN PS WITHOUT TCSC

Line	Branch		Power flow	Power losses
	From node	To node	[MW]	[MW]
1	1	2	440,144	2,439
2	2	3	137,329	0,376
3	3	4	-162,671	0,258
4	4	5	-462,929	2,150
5	5	6	-179,046	0,435
6	6	7	395,567	1,687
7	7	8	306,441	0,964
8	5	7	213,967	0,443
9	8	9	6,424	0,017
10	9	10	-294,649	1,073
11	1	10	172,181	0,829
12	6	10	423,297	1,655
Total power losses				12,325

TABLE III
POWER FLOW AND TOTAL POWER LOSSES IN PS WITH TCSC

Line	Branch		Powe flow	Power losses
	From node	To node	[MW]	[MW]
1	1	2	419,06796	2,141
2	2	3	116,92689	0,262
3	3	4	-183,3352	0,318
4	4	5	-483,6534	2,353
5	5	6	-190,8222	0,498
6	6	7	398,52498	1,678
7	7	8	301,27154	0,896
8	5	7	204,81604	0,391
9	8	9	0,3755584	0,021
10	9	10	-299,6452	1,094
11	1	10	193,04125	0,960
12	6	10	410,15524	1,497
Total power losses				12,109

In this case, we are able to reduce total power losses by 1,75 % at 60 % line compensation. TCSC installed in other lines, was found only a very small reduction in total power losses. It should be noted that in this case we control power flow in terms of minimizing losses. For such control do not fall power losses in all lines, but total power losses are lower, as can be seen from tables 2 and 3.

V. CONCLUSION

As shown in this article using TCSC, we are able not only to regulate active power flow, but also reduce the total power losses in the PS. Use of equipment for only this purpose in the future is questionable, because the financial cost of these devices are high. It is all on an individual assessment of the suitability of the location of TCSC in PS.

ACKNOWLEDGMENT

This work was supported by Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences under the contract No. 1/0166/10 and by Slovak Research and Development Agency under the contract No. APVV-0385-07 and No. SK-BG-0010-08.

REFERENCES

- [1] Johns, A. T., TER-GAZARIAN, A., WARNE, F., "Flexible ac transmission systems (FACTS)," The Institution of Electrical Engineers, 1999, 592 pp, ISBN 0-85296-771-3.
- [2] MATHUR, R. M., VARMA, R. K., "Thyristor-based FACTS controllers for electrical transmission systems," Institute of Electrical and Electronic Engineers, 2002, 493 pp, ISBN 0-471-20643-1.
- [3] ACHA, E., FUERTE-ESQUIVEL, C. R., AMBRIZ-PÉREZ, H., ANGELES-CAMACHO, C., "FACTS Modelling and Simulation in Power Networks," John Wiley & Sons, 2004, 421 pp, ISBN 0-470-85271-2.
- [4] HINGORANI, G. N., GYUGYI, L., "Understanding FACTS. Concepts and technology of Flexible AC transmission Systems," New York: IEEE Press, 2000, 432 pp, ISBN 0-7803-3455-8

Author's index

A

ADAMONDY Peter 140

B

BAČÍKOVÁ Michaela 19

BAČKO Martin 340, 369, 424

BALÁŽ Anton 113, 129

BALOGH Tibor 309

BATMEND Mišél 148

BÉREŠ Tomáš 352

BLICHA Radovan 133

BODOR Marcel 388

BORDUG Dmytro 369

BRATASH Oksana 340

C

CABÚK Pavol 316

CIPOV Vladimír 23

COPÁK Marek 23

CYMBALÁK Dávid 185, 219

Č

ČAJKOVSKÝ Marek 162

ČONKA Zsolt 355

ČOPÍK Matej 72, 212

D

DANKOVÁ Eva 124

DEMETER Dominik 392

DUDIAK Jozef 279

DUFALA Marek 103

DUPÁK Denis 133

DURANKA Peter 267, 372

E

ENNERT Michal 143, 162

F

FABRI Daniel 333

FANFARA Peter 233

FÁBRI Daniel 291

G

GALLO Patrik 208

GAŠPAR Vladimír 158

GERMAN-SOBEK Martin 255,
259

GLOD Lukáš 365, 421

GODLA Marek 250, 291, 333

GONTKOVIČ Daniel 222

H

HALAJ Marián 362

HALUPKA Ivan 68, 84

HARASTHY Tomáš 299, 306

HAVRILOVÁ Cecília 158

HOCKO Pavol 279, 427, 431

HOŠÁK Rastislav 225

HRÁŠOK Tomáš 243

HRIC Matúš 271, 302

HRINKO Marián 396

HROZEK František 107

CH

CHODAREV Sergej 31, 60

I

ILKOVIČ Ján 72, 177

ISTOMINA Nataliia 337

IVANČÁK Peter 107, 143

J

JADLOVSKÁ Slávka 38

JAČIŠIN Štefan 34

JAKUBČÁK Roman 414, 435

JENČIK Marián 96

JURČIŠIN Michal 316

K

KALINOV Andrii 286, 417

KALAVSKÝ Michal 348, 352, 375

KAPA Martin 76

KAROL Tomáš 158

KATIN Matúš 414, 435

KELTIKA Marián 166

KHREBTOVA Oksana 275

KIRÁLY Jozef 255, 259

KLIMEK Ivan 121, 162, 166

KOKOŠKA Rastislav 382

KOŠICKÝ Tomáš 344

KOVÁČ Ondrej 154, 295, 319

KRAVČÍK Michal 282

KRAVETZ Oleksii 247

KRIŠTOF Vladimír 359

KUŠNÍR Stanislav 355, 359

KYSLAN Karol 271, 302

L

LACIŇÁK Lukáš 117, 216

LACKO Milan 388

LAKATOŠ Dominik 92

LESHCHUK Oleksii 275

LIPTAJ Martin 243, 322

LIPTÁK Jozef 250, 291, 333

LUKÁČ Gabriel 52

LUKÁČOVÁ Alexandra 64

Ľ

ĽALOVÁ Martina 216

M

MACKO Pavol 151

MACH Marián 88

MAMCHUR Dmytro 286, 417

MIHAL Roman 117

MIKULÁK Martin 148

MIŠENČÍK Pavol 299, 306

MIŽENKO Ladislav 181

MOLČAN Anton 56

MOLNÁR Ján 286, 337, 424

N

NGUYEN Peter 313, 403

NOSÁL Milan 42, 204

NOVÁK Marek 76

NOVÁK Matúš 427, 431

O

OCILKA Matúš 247, 337

OGAR Vita 340

P

PALA Martin 181

PALOVÁ Dana 169

PAPCUN Peter 140, 177, 212

PAVLÍK Marek 329, 396, 407

PAVLÍK Miloš 225

PÁSTOR Marek 239

PEKÁR Adrián 103, 124, 233

PERDULAK Ján 275, 411

PEŤKO Ivan 196

PIETRIKOVÁ Emília 31, 68, 84

R

RAKOCI František 400

REPKA Martin 99

RINÍK Vojtech 121

ROČKAI Viliam 88

ROMASHIHIN Yuriy 424

RUMAN Kornel 295, 319

S

SEKERÁK Martin 250

SERBÁK Vladimír 136

SHUTKA Oleksander 247

SMOLÁR Peter 27, 192

Š

ŠESTÁK Kristián 228

ŠTOFA Ján 169

ŠUHAJOVÁ Viktória 326

ŠUSTER Peter 188

T

TATARKO Matúš 379

TUŠANOVÁ Adela 110

U

UHRÍNOVÁ Magdaléna 267, 372

URDZÍK Daniel 263

V

VACEK Marek 313, 403

VALISKA Ján 154, 295

VALO Matúš 200

VARGA Martin 173

VASZIOVÁ Gabriela 365, 421

VAVREK Jozef 80

VEHEC Igor 319

VINCE Tibor 369, 411, 417

VIRČÍKOVÁ Mária 27, 192

VISZLAY Peter 15

VOZÁRIKOVÁ Eva 48

Z

ZACHEPA Iurii 411

ZBOJOVSKÝ Ján 329, 407

Ž

ŽIGA Matej 322

**12th Scientific Conference of Young Researchers
of Faculty of Electrical Engineering and Informatics
Technical University of Košice**

Proceedings from Conference

Published: Faculty of Electrical Engineering and Informatics
Technical University of Košice
I. Edition, 440 pages

Available: http://web.tuke.sk/scyr/data/templates/Proceedings_2012.pdf

Editors: Prof. Ing. Alena Pietriková, CSc.
Ing. Dominik Demeter
Ing. Milan Nosál

ISBN 978-80-553-0943-9