

WOFEX 2013

proceedings of the 11th annual workshop

Faculty of Electrical Engineering and Computer Science,
VŠB – Technical University of Ostrava

ISBN 978-80-248-3073-5

WOFEX 2013 wofex

<http://wofex.vsb.cz/2013/>

11th annual workshop,
Ostrava, 10th – 11th September 2013
Proceedings of papers

Organized by

VŠB – Technical University of Ostrava
Faculty of Electrical Engineering and Computer Science



WOFEX 2013

© Michal Krátký, Jiří Dvorský, Pavel Moravec, editors

ISBN 978-80-248-3073-5

This work is subject to copyright. All rights reserved. Reproduction or publication of this material, even partial, is allowed only with the editors' permission.

Page count: 565
Impression: 200
Edition: 1st
First published: 2013

This proceedings was typeset by PDFL^AT_EX.

Published by Faculty of Electrical Engineering and Computer Science, VŠB – Technical University of Ostrava

Preface

The workshop WOFEX 2013 (PhD workshop of Faculty of Electrical Engineering and Computer Science) was held on September 10th – 11th September 2013 at the VŠB – Technical University of Ostrava. The workshop offers an opportunity for students to meet and share their research experiences, to discover commonalities in research and studentship, and to foster a collaborative environment for joint problem solving. PhD students are encouraged to attend in order to ensure a broad, unconfined discussion. In that view, this workshop is intended for students and researchers of this faculty offering opportunities to meet new colleagues.

This book of proceedings includes 85 papers of faculty PhD students and 6 papers of external authors. The proceedings of WOFEX 2013 are also available at WOFEX Web site <http://wofex.vsb.cz/2013/>. I would like to thank the authors and the Organizing Committee from Department of Computer Science, namely Jiří Dvorský and Pavel Moravec, for their arduous editing work.

September 2013

Michal Krátký
Program Committee Chair
WOFEX 2013

Organization

Program Committee

Chair:

Michal Krátký (VŠB – Technical University of Ostrava)

Members:

Pavel Brandštetter (VŠB – Technical University of Ostrava)

Radim Briš (VŠB – Technical University of Ostrava)

Zdeněk Dostál (VŠB – Technical University of Ostrava)

Petr Chlebiš (VŠB – Technical University of Ostrava)

Stanislav Rusek (VŠB – Technical University of Ostrava)

Václav Snášel (VŠB – Technical University of Ostrava)

Vladimír Vašínek (VŠB – Technical University of Ostrava)

Organizing Committee

Jiří Dvorský (VŠB – Technical University of Ostrava)

Pavel Moravec (VŠB – Technical University of Ostrava)

Kateřina Kašparová (VŠB – Technical University of Ostrava)

Workshop Location:

Campus of VŠB – Technical University of Ostrava

17. listopadu 15, 708 33 Ostrava–Poruba, Czech Republic

10th – 11th September 2013

<http://wofex.vsb.cz/2013/>

Table of Contents

WOFEX Workshop

Study Programme P2649 Electrical Engineering

Electronics (2612V015)

Microprocessor Control Voltage Inverter	1
<i>Martin Borski</i>	
Realization of Microprocessor Control One and Three-Phase Thyristor Rectifier	6
<i>Jiří Špička</i>	
Thermal field in the heat sink KL 494	12
<i>Roman Šprlák, Josef Opluštíl, Daniel Kalvar, and Miroslav Uchoč</i>	
Analog Electronic Circuits with Functional Blocks Based on Current Conveyors	18
<i>Radek Šulák</i>	
Simulation of the Three Phase DC/DC Converter used for Designing and Testing of Traction Battery Models	24
<i>Martin Tvrdon and Ales Havel</i>	

Study Programme P1807 Computer Science, Communication Technology and Applied Mathematics

Computer Science (1801V001)

Compression and Multiple Range Query in Tree Data Structures	29
<i>Václav Bašniar</i>	
Processing of Range Query in R-tree Using CPU and GPU	35
<i>Pavel Bednář</i>	
Polyaniline – constant electrical resistance	41
<i>Jiří Bocheza and Stanislav Hejduk</i>	
Energy Based Descriptors and their Application for Object Detection	46
<i>Radovan Fusek, Eduard Sojka, Karel Mozdřeň, and Milan Šurkala</i>	
Image Segmentation via Graph Cuts using Residual Graph	52
<i>Michael Holuša</i>	
Indoor localization system using WiFi fingerprinting and built-in inertial sensors	58
<i>Jakub Jirka</i>	

Software Process Development for Healthcare Information Systems Based on Formal Methods	64
<i>Michael A. Košinár</i>	
QuickXDB: Native XML Database	70
<i>Petr Lukáš</i>	
Influence of partiality on procedural isomorphism	76
<i>Jakub Macek and Lukáš Vích</i>	
Debugging & Profiling of MPI Applications Generated by Kaira	82
<i>Ondřej Meca and Martin Šurkovský</i>	
Background Subtraction using a Layered RC Circuit Model	88
<i>Karel Mozdřeň, Eduard Sojka, Radovan Fusek, and Milan Šurkala</i>	
An Application of Process Mining to Real Business Processes	94
<i>Jakub Stolfa</i>	
Hierarchical Layered Mean-Shift Methods	100
<i>Milan Šurkala, Karel Mozdřeň, and Radovan Fusek</i>	
Introduction to financial markets trading	106
<i>Jiri Urbasek</i>	

External Authors

An analysis of an object's parameters by means of image processing	112
<i>Rafał Gasz</i>	
The Use of Absolute Vibrations for On-line Monitoring of Industrial Electric Motors	118
<i>Marek Klimowski</i>	
Analysis of Iron and Copper Losses in Permanent Magnet Synchronous Motor	124
<i>Piotr Mynarek</i>	
Evaluation of algorithms of SIFT, BRIEF, FREAK descriptors, and their effectiveness when comparing images undergoing geometric translations	128
<i>Marcin Niedźwiecki</i>	
Implementation of the Plume Rise Procedure in the Massively Parallel Architecture	134
<i>Tomasz Przybyła</i>	
Predictive algorithms for three-wheeled mobile robot control	138
<i>Tomasz Pyka</i>	

Study Programme P2649 Electrical Engineering

Electrical Power Engineering (3907V001)

Calculation of Locomotive Tractive Power During the Starting Phase	147
<i>Maroš Ďurica</i>	
Comparison of photometric properties of the projection screen and interactive whiteboard with EN 12464-1 at classroom	153
<i>Barbara Helštýnová, Karel Sokanský, and Tomáš Novák</i>	
Technical condition and importance of line for 110 kV power lines	160
<i>Vít Houdek</i>	
Annual Energy Balance of Biogas Station	166
<i>Jiří Janša</i>	
Active charging stations for electric cars	172
<i>Oliver Marcínčin and Zdeněk Medvec</i>	
Power quality measurement	178
<i>Michal Reguľa</i>	
Complaints about the quality of electricity in Northern Moravia	184
<i>Petr Rozehnal</i>	
Photovoltaic Power Plant, Extreme Thresholds of Production	190
<i>Martin Smoček and Zdeněk Hradílek</i>	
Monitoring of Hybrid system using LabView – Next step in development .	196
<i>Jindřich Stuchlý</i>	
Modelling of a DVR model based on DQ transformation control technique in MATLAB/Simulink	201
<i>Dominik Szabó</i>	
Design, realization and analysis-methods for measurement heat pump energy balance	207
<i>Jan Šrámek</i>	
Measures to achieve energy savings of lighting systems using daylight	213
<i>Jan Šumpich, Zbyněk Carbol, and Karel Sokanský</i>	
Analysis of power grid including energy storage station using EMTP-ATP	219
<i>Marián Uher</i>	
Energy rating of hydrogen system at VSB-TUO	225
<i>Jan Vaculík</i>	

Off-grid hybrid power system	231
<i>Jakub Vramba</i>	

Electrical Machines, Apparatus and Drives (2642V004)

Sensorless Control of Induction Motor Using State Observers	237
<i>Jiří Friedrich</i>	

The Use of the General Theory of Electrical Machines for Modeling of Induction Motor in Matlab Simulink	243
<i>Stanislav Rusnok, Pavel Sobota, and Pavel Svoboda</i>	

Identification and Optimization of Parameters of Synchronous Machines with PM	249
<i>Ladislav Slíva</i>	

Sensorless Control of Induction Motor Using Injection Method	255
<i>Tomáš Verner</i>	

Technical Cybernetics (2612V045)

Safe HMI and Car to Car Communication	261
<i>Radim Hercík</i>	

Power Analyzer for Measurement of Efficiency of the Frequency Converter	266
<i>Marek Horinek and Petr Bilík</i>	

E-Bike Testing Station	272
<i>Jaromír Konečný</i>	

Hybrid Power System Power Flow Analysis	277
<i>Jakub Maňas</i>	

Knowledge-based Monitoring and Re-adaptation of PID Controllers	282
<i>Jana Nowaková and Miroslav Pokorný</i>	

Remote control of laboratory task using wireless technologies and operating system Android	288
<i>Vít Otevřel and Zdeněk Slanina</i>	

Verification methodology for noninvasive continuous measurement of blood pressure	294
<i>Lukas Peter and Martin Cerny</i>	

Testing gapless measurement of PQA according to IEC 62586-2	300
<i>Petr Pěťvaldský</i>	

System for Automatic Loading under Operating System QNX	306
<i>Roman Slabý and Martin Stankuš</i>	

The SSVEP Detection Using the Canonical Correlation Analysis and the Continuous Wavelet Transform	311
<i>Maros Smondrk, Michal Vozda, Jana Krohova and Barbora Hrvolova</i>	
RFID Expert Database	317
<i>Lucie Švecová</i>	
Using Mathematical Background for Determining the Temperature Distributions and Corresponding Parameters inside a Rock Massif	321
<i>Petr Vojčínák</i>	
Study Programme P1807 Computer Science, Communication Technology and Applied Mathematics	
Computational and Applied Mathematics (1103V036)	
Risk Modeling of Accidental Events and Responses in Process Industries .	327
<i>Ondřej Grunt</i>	
Efficient Implementation of the TFETI Coarse Problem Solution in the FLLOP library	333
<i>Václav Hapla</i>	
Comparison of laparoscopic and open surgery techniques in liver surgery .	339
<i>Martin Hasal</i>	
Identification problems with a priori given material interfaces	345
<i>Rostislav Hrtus, Jaroslav Haslinger, and Radim Blaheta</i>	
Weibull Proportional Hazards Regression Model for Patients with Surgery of Colon	351
<i>Kateřina Janurová</i>	
Tangential Fields in Mathematical Modelling of Optical Diffraction Problems	357
<i>Jiří Krček and Jaroslav Vlček</i>	
Parallel implementation of fast boundary element method	363
<i>Michal Merta</i>	
Semi-Smooth Newton Method for Contact Problems with Tresca and Coulomb Friction	369
<i>Kristina Motyčková and Radek Kučera</i>	
Parallel Hybrid Mesh Multiplication	375
<i>Aleš Ronovský</i>	
Applied graph theory in Markov chains reliability analysis	381
<i>Pavel Skalný</i>	

Comparison of numerical codes for diagonalization of Hermitian matrices 386
Martin Stachoň

Elastic image registration with local refinement 392
Alena Vasatova and Aleš Ronovský

Shape Optimization Based on the Shape Calculus and the Boundary
Element Method 398
Jan Zapletal

Communication Technology (2601V018)

The convergence of OpenBTS and OpenIMSCore projects 404
Marcel Fajkus, Jakub Cubik, and Stanislav Kepak

Visible Light Communication with Four Wavelengths Division Access 409
Stanislav Hejduk, Jiri Bocheza, Karel Witas, and Artem Ganiyev

Measurement of homogeneity of the microwave emitter 415
Jakub Jaros, Andrej Liner, and Martin Papes

Integration of voice services into corporate information systems 421
*Martin Mikulec, Jan Rozhon, Filip Řezáč, Jiri Vychodil, and
Miroslav Vozňák*

Recognizing emotions from human speech using 2-D neural classifier
and influence the selection of input parameters on it's accuracy 428
Pavol Partila, Karel Tomala, and Jakub Safarik

Study of Electromagnetic Compatibility between Railway Infrastructure
and Train Interface 434
Phan Thi Thanh Thao and Lubomír Ivánek

Recovery of LED Diode Optical Spectrum 441
Jan Vitásek and Jan Látal

Study Programme P1807 Computer Science, Communication Technology and Applied Mathematics

Computer Science (1801V001)

Similarity detection of Image using Vector quantization and Compression	453
<i>Petr Berek, Michal Průpek, and Václav Snášel</i>	
Source localization for EEG patterns relevant to motor imagery BCI control	459
<i>Pavel Bobrov, Alexander Frolov, Dušan Húsek, and Jaroslav Tintěra</i>	
Common Tensor Discriminant Analysis: A Massively Parallel Implementation with Application to Human EEG Wave Recognition	465
<i>Pavel Dohnálek</i>	
Microsoft Kinect and persons with disabilities	471
<i>Jaroslav Hořejší and Adam Kašpar</i>	
Visualization of Large Graphs Using GPU Computing	477
<i>Tomáš Jeřowicz, Miloš Kudělka, Jan Platoš, and Václav Snášel</i>	
Visualization of data using Fibonacci spiral	483
<i>Jonáš Krutil, Miloš Kudělka, and Václav Snášel</i>	
Similarity in EEG data using Lempel-Ziv Complexity	489
<i>Michal Průpek, Ibrahim Salem Jahan, and Václav Snášel</i>	
Introduction to de novo DNA assembly algorithms	495
<i>Daniel Robenek, Jan Platoš, and Václav Snášel</i>	
Stock price prediction based on particle swarm optimization	501
<i>Lukáš Zaorálek and Tomáš Buriánek</i>	

**Study Programme P1807 Computer Science,
Communication Technology and Applied
Mathematics**

Computer Science (1801V001)

Finding Evolution and Measuring of Stability of Co-Authors Communities Formed by Terms on DBLP	511
<i>Alisa Babskova</i>	
Comparison of graph compression algorithms	517
<i>Patrik Dubec and Jan Plucar</i>	
Solving Steel Alloying Using Differential Evolution and SOMA	523
<i>Michal Holis, Lenka Skanderova, Martin Placek, Jiri Dvorsky, and Ivan Zelinka</i>	
Comparison of Bio-inspired and Mathematical Optimization Methods on a Graph Drawing Problem	529
<i>Lukas Rapant</i>	
Architectures and Features of Modern Honey pots	535
<i>Daniel Střibný and Zbyněk Složil</i>	
Growing Neural Gas – A Parallel Approach	541
<i>Lukáš Vojáček</i>	

Author Index

Microprocessor Control Voltage Inverter

Martin Borski

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
`martin.borski@vsb.cz`

Abstract. This paper describes the design and realization of voltage inverter with microprocessor control. At first is shown the completely connection with the description of individual components. And then are performed several measurements. The created model will be used for innovative laboratory exercises and study purposes in the Department of Electronics. The converter includes conventional modules with IGBT (Isolated Gate Bipolar transistor) controlled by a modern control system. The processor in the control unit works in real – time algorithm.

Keywords: converter; voltage inverter; power modules; IGBT; IGBT driver; capacitors battery; control unit

1 Introduction

The development of modern control units with fast signal processor causes advancement in the control of the power switching devices. Thanks to these control units we receive control with optimal results. Power switching parts are constantly being developed by using new technologies, too. Their parameters increase and they need a faster and powerful control system. This also applies to a device IGBT. These item's excellent frequency properties are mostly used in power converters.

2 Power Voltage Inverter

The inverter can be powered from a DC source as well as from three-phase network, because a rectifier is connected. The fuses are included in the input for protection of semiconductor components when an undesirable effects (short-circuit, overload). Capacitor in the DC link is a source smoothed DC voltage for the inverter and load for the rectifier at the same time. Sensors on the input and output measure values and send it to the control unit. The control unit communicates with the host computer and controls the switching of individual transistors.

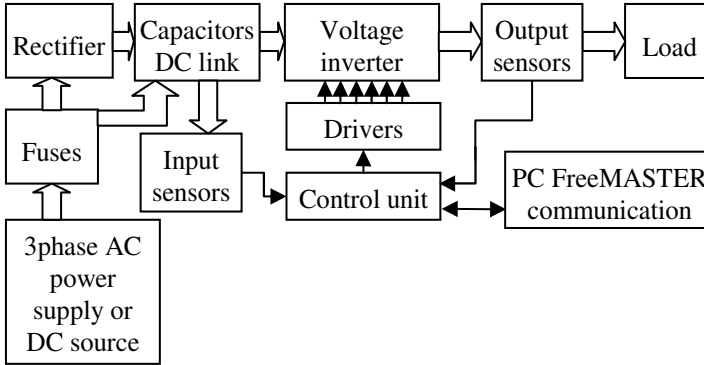


Fig. 1. Controlled voltage inverter block diagram

2.1 Description of a Block Diagram

Protection. The converter is protected by five cylindrical fuses with marking PV51432AgR. These fuses serve for turning-off the nominal current 32 A. These fuses are inserted into the fuse switch-disconnectors OPV14.

Rectifier. SKD50/12A3 is used as the power bridge rectifier. This is a three phase diode-rectifier in an isolated metal case with screw terminals.

Capacitors battery. In a converter's DC-link, there is a capacitors battery located with an equivalent capacity 3.3 mF. This value is achieved by six electrolytic capacitors type K01 2.2 mF. Balancing resistors 27 kΩ are connected in parallel with the capacitors.

Power IGBT modules. Inverter's power bridge is created by IGBT transistors with inverse diodes. Three modules of type SKM75GB123D by Semikron are used in this three-phase inverter design. In one case (Semitrans 2) there are two transistors. The maximum range of voltage between collector-emitter is 1200 V and continual current 75 A. As close as possible to these IGBT modules are located snubber capacitors which reduce the inductance between the switch and the capacitor.

IGBT drivers. Drivers by Semikron of type SKHI22AH4R are used in this application. This is a hybrid dual IGBT driver in a case SEMIDRIVER. The Driver is isolated by transformers and it has an integrated circuit for protecting the power supply under voltage (13 V) and short circuit.

Voltage and current sensors. In the converter, there are generally eight sensors by Swiss company LEM. There are four LV25-P voltage sensors and four LAH 25-NP current sensors. One of the voltage and current sensors measures the voltage and current of the DC-link. The other sensors measure voltage and current on output phases.

Control system. Microcontroller Freescale 56F8037 is used for switching IGBT transistors.

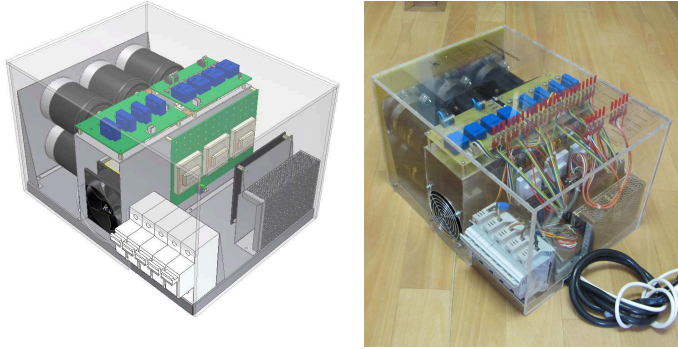


Fig. 2. 3D inverter model and completely assembled inverter

3 The Construction of Inverter

With the help of 3D CAD, the converter design enables the exact layout of components and their production if needed. The isolation distance and the cable placement must be respected in the design.

A complete 3D CAD model of the voltage inverter without cables is in the picture (Fig. 2).

Aluminium heat sink L 280 type 35.2 is in the middle part. The fan is dissipated the heat from the heat sink. So the air is circulated and forced convection is secured. A Power rectifier module and IGBT modules are strongly mounted by screws on aluminum heat sink. The distance between the modules is 15 mm.

In the picture (Fig. 2), there is a complete converter with a laminate cover. The control system is placed on an independent metal shield due to isolation distance and EMC (Electromagnetic Compatibility). In front of the converter, there are a power source and fuse switch-disconnectors for protection leads. Capacitors are attached by ringing clips in a horizontal position on the back side of the converter.

4 Experimental Results

The individual parts were tested before the first starting operation of the inverter (control system, drivers, measuring sensors, etc.). The measurement was performed on a completely compiled converter.

The inverter was connected to an asynchronous motor and was tested under load (Fig. 3). Voltage and current waveforms were measured at the output of the inverter.

The measuring points lead out of each IGBT to the cover. There are a collector, an emitter, a gate and an output of the driver before a resistor R_{Gon} . This makes it possible, for example to measure the pulse graph is between the gate and the emitter transistors in one phase (Fig. 4) or the dead time is measured

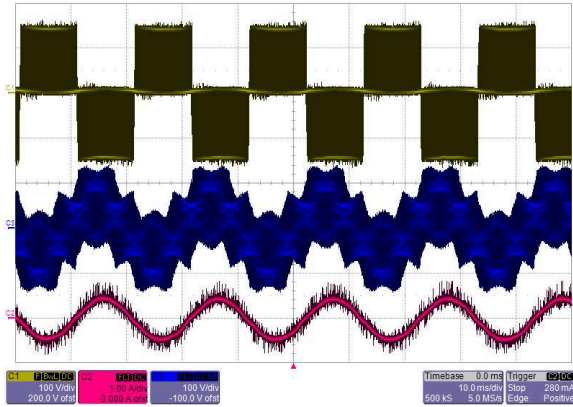


Fig. 3. Graph line output voltage (yellow), phase voltage (blue) and current (red)

on transistors in one branch of the inverter (Fig. 5). The signals are brought out from individual sensors. All of these variables can be measured without the need of the protective cover being removed.

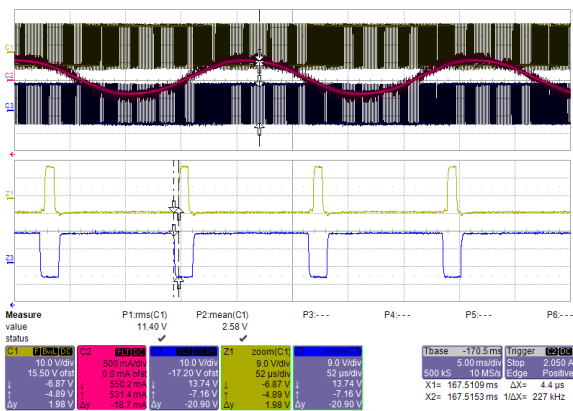


Fig. 4. Switching pulses on transistors in one phase

5 Conclusion

With the use of innovative processes and modern parts a laboratory model of a voltage inverter was created. Its functionality was proved by measurements on its individual parts as well as on the device as such. Supposed outputs were reached. It will be possible to make different measurements on the inverter and innovating already set assessments. It will also be possible to assess bachelor and

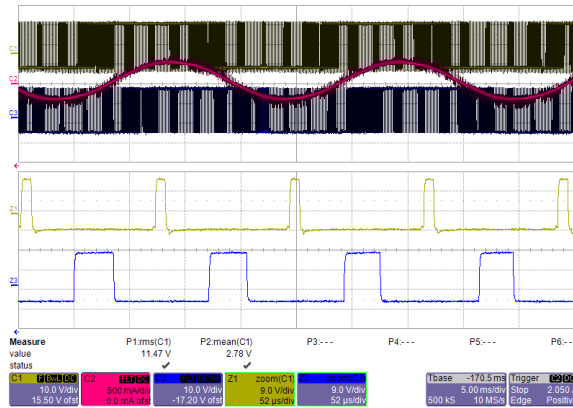


Fig. 5. Dead time on transistors in one branch inverter

diploma thesis based on the inverter. It is necessary to suggest other means of driving and optimization of the operation.

Acknowledgment

The research described in this paper was supported by these projects:

- SGS project SP2013/80 – Utilizing of Modern SiC Components for MF Power Semiconductor Converters.

This research was supervised by prof. Ing. Petr Chlebiš, CSc. from Department of Electronics, VSB - TU Ostrava.

References

1. Havel, Ales, Petr Vaculik and David Slivka. Efficient Propulsion Structure with Anaxial Flux Rotary Converter for HEV Drive Unit. Advances in Electrical and Electronic Engineering. September 2011, vol. 9, iss. 3,. ISSN: 13361376.
2. Slivka, David, Petr Palacky, Petr Vaculik and Ales Havel. Electric Vehicle Control units Communication. Advances in Electrical and Electronic Engineering. March 2012, vol. 10, iss. 1, p. 17-21,. ISSN:13361376.
3. Palacky, Petr, Martin Sobek, Petr Hudecek, David Slivka and Rostislav Naivert. A Multipurpose Control Unit Based on Freescale 56F8037 with Implemented Vector Control of an Induction Motor. Advances in Applied Electronics. September 2010, p. 241-244, ISSN:1803-7232.
4. Neborak, Ivo, Petr Skotnica. Electric Car with Induction Motor Supplied by Voltage Inverter. Advances in Electric Power Engineering. May 2010, p. 823-826, ISBN:978-80-214-4094-4.

Realization of Microprocessor Control One and Three-Phase Thyristor Rectifier

Jiří Špička

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jiri.spicka@vsb.cz

Abstract. This article describes the design, simulation and realization of the thyristor controlled rectifier with the microprocessor control unit. The created power semiconductor converter is used for teaching in the Department of Electronics, VŠB - Technical University of Ostrava. The power converter is based on 56F8037 digital signal controller. The realized controlled rectifier and experimental results are added to end of the article.

Keywords: controlled rectifier, power converter, thyristor, control unit, digital signal controller 56F8037

1 Introduction

The controlled rectifiers belongs to a group of the inverters with external commutation. They converts AC electrical power to DC power. They are used to control a motors (such as soft-starters AC motors), to control the lighting, temperature control of the furnace or temperature control for a chemical processes, etc.

An experimental measuring station with the controlled rectifier is used to describe the principles of uncontrolled and controlled rectifiers for various applications. The measurements on the power rectifier describe the transfer of energy flow from the power supply to the load when changing the required parameters, commutation in the rectifier, reverse effects back to the power supply at different parameters of the load. A harmonic analysis of the phase currents is also implemented on the experimental measuring station with the controlled rectifier.

2 Description of parts of the power converter

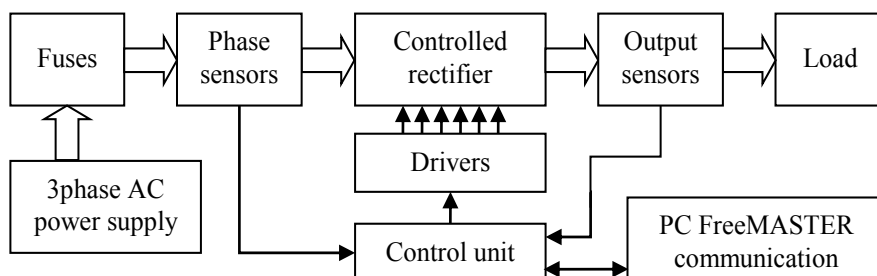


Fig. 1. Block diagram of the controlled rectifier

2.1 Description of the block diagram

Controlled rectifier. The controlled rectifier is constructed in a three-phase bridge. A two-pulse or six-pulse rectifier is selected by appropriate control. The rectifier is composed of a three modules SKKT. Each module contains a pair of the thyristors. The thyristor bridge is mounted to the heat sink type L140 34.2. The heat sink is actively cooled.

Fuse. The protection of the experimental station with the controlled rectifier is divided into two parts. The first part is a software protection and the second part is a fuse protection. The rectifier is protected against high phase currents and output short circuit by the software. The fuses protect the rectifier against high phase currents.

Sensors. The phase and output signals are scanned by sensors made by LEM. The current sensors are indicated as LAH 25-NP and voltage sensors are indicated as LV25-P. The current signal of the sensor carries information about a shape of scanned variable. This value is processed by analog-to-digital converter.

Control unit. The control unit is a microprocessor board with digital signal controller 56F8037 by Freescale. Eight analog-to-digital converters, two digital-to-analog converters and three programmable interrupt timers are needed to drive the thyristor controlled rectifier.

Drivers. The energizing circuit supplied power for switching the thyristor and these circuits provide galvanic isolation of the control part from the power part of the rectifier. The galvanic isolation is realized by an optocouplers (for the control unit) and by DC/DC converters (power for the energizing circuits).

FreeMASTER. The experimental measuring station is controlled by program FreeMASTER. The control unit communicates with the FreeMASTER through programming and communication interface JTAG.

2.2 Parameters of the controlled rectifier

Phase voltage $U_{RMS} = 50 - 230V / 50 - 60Hz$

Phase current max $I_{RMS} = 25A$ (by sensors)

Fuse current protection $I_{RMS} = 15A$

Software current protection $I = 12A$

Dimensions 315 x 210 x 200 mm

3 Simulation and design

The control section and the power section of the controlled rectifier are simulated by OrCAD - PSpice. A dimensioning of the converter in terms of heat loss and calculations of electrical variables are calculated by MATLAB software.

3.1 Simulation

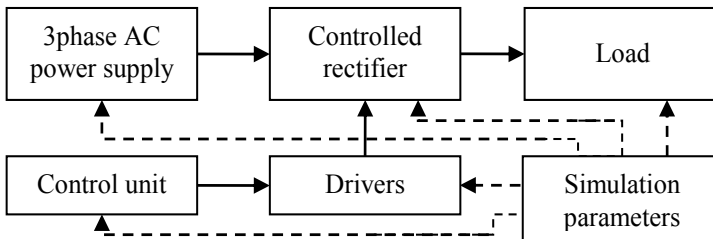


Fig. 2. Simulation model of the controlled rectifier

The simulation scheme is composed of blocks in the hierarchical structure of the program OrCAD. The simulation environment is designed to resemble the real conditions. In the simulation is developed the control electronics of the rectifier according to real design, implementation impedance of the power supply, refining PSpice models of the thyristor modules etc. The simulation is controlled by a global parameters. These parameters change an internal impedance of the power supply (such as increasing input inductance), control angle α , generated and adjusted the widths of the control pulses. Type of the load is also controlled by the global parameters.

The simulation of the controlled rectifier led to a number of simulation results that indicate the proper design of the rectifier.

3.2 Design cooling of the controlled rectifier

When designing a suitable cooling of the controlled rectifier is based on the assumption that the power dissipation of the rectifier is mainly due to the size of a forward current of the thyristor (1). Cooling of the thyristor controlled rectifier is designed to limit parameters of the thyristor modules. The thermal models are constructed for both two-pulse, and for the six-pulse controlled rectifier bridge.

Thyristor modules datasheet specifies the maximum mean value a forward current of thyristor I_{TAV} . The maximum value of the current flowing through the thyristor is given by the maximum supply voltage (phase-to-phase voltage) and minimum load resistor.

$$P_{FAV} = U_{TO} \cdot I_{TAV} + r_r \cdot I_{RMS}^2 \quad (1)$$

The power dissipation of the module P_{FAVM} is double that the power dissipation of the thyristor because each thyristor current leads one half period in the two-pulse rectifier.

$$\Delta T = P_{FAVM} \cdot \sum R_{th} \quad (2)$$

In a similar way, the assembled thermal model of the six-pulse rectifier. The difference is that the power dissipation of the thyristors (1) is given by other passage of the currents than the thyristors in the two-pulse rectifier bridge.

For the thyristor controlled rectifier is chosen the heat sink with active cooling model L140 type 34.2. with the thermal transient resistance of the heat sink assembly $R_{thh-a} = 0,185\text{K/W}$.

3.3 Design fusing of the controlled rectifier

Design of the fuse protection of the thyristor controlled rectifier is based on the current waveforms of the phases, calculations of their effective values depending on the mean values of current waveforms of the thyristors. As in the previous chapter the design calculated with the two-pulse and six-pulse rectifier bridge.

Used current sensors and cable management, do not allow maximum workload of the thyristor modules. Maximum size of effective value of current flowing through the sensor $I_{RMS} = 25\text{A}$.

4 3D CAD design and construction

The parts and overall model of the controlled rectifier are designed in 3D CAD system Autodesk Inventor. 3D CAD model of the controlled rectifier was made especially for the visualization of a suitable arrangement of the individual components.

The controlled rectifier is covered by a plexiglass. Appropriate choice of the cover allows insight into the internal structure of the power converter and also serves as a protection against accidental contact with live parts.

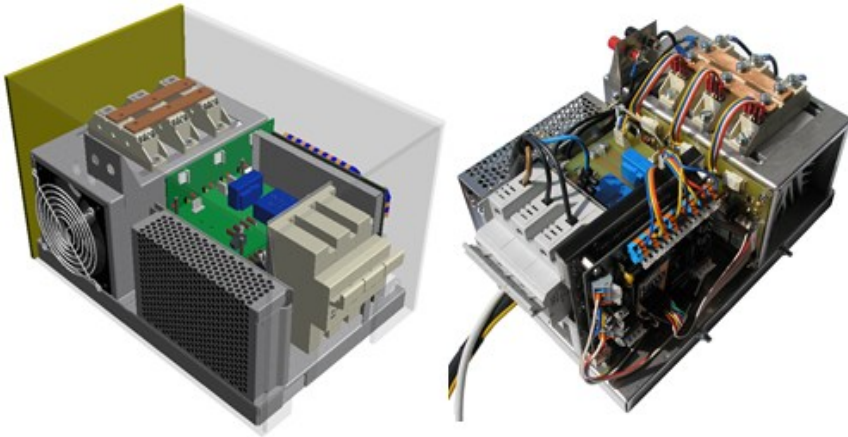


Fig. 3. 3D CAD model and realized controlled rectifier

Construction of the controlled rectifier based on 3D CAD model of power converter. An important part of the design and the implementation was reviving control electronics, control unit, drivers, input and output sensors. Another part was a proposal of the power supply of the control electronics and cooperation control electronics with power section of the thyristor controlled rectifier.

5 Experimental results

The first measurements of the rectifier were aimed to verify its correct operation. For example, the functionality of the drivers or sensing power magnitudes by sensors.

The experimental result of two-pulse rectifier with RL load and the inclusion of the input choke on the power supply. The measurement shows the shapes of the output voltage (yellow) and current (red) in the power supply voltage (blue) and backward influences of the rectifier on the power supply when the control angle $\alpha = 55^\circ$.

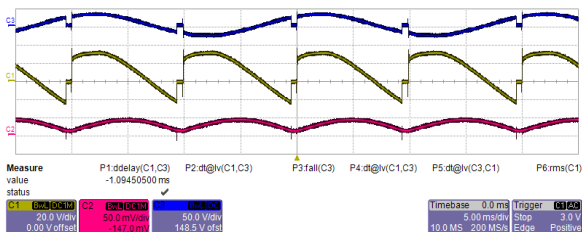


Fig. 4. Voltage and current waveforms of the two-pulse rectifier

Another experimental result is the measurement commutation voltage drop under high input impedance of the power supply. The output voltage of the six-pulse rectifier and the couple phase currents are showed on the figure 5. The Graph show the commutation time.

Thermal field in the heat sink KL 494

Roman Šprlák, Josef Opluštil, Daniel Kalvar, and Miroslav Uchoč

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
roman.sprlak@vsb.cz

Abstract: This article deals with simulation of thermal field in the KL-494 liquid cooled heat sink from Seifer Electronic. This heat sink can be used in frequency converter and this converter can be used in locomotive. The 3D model of heat sink was created in 3D CAD program. The power modules from Semicon are placed on the heat sink. Type of power semiconductor module is SEMiX854GB176HDs. This heat sink is 630 mm long and it is made from aluminum. The main idea of this paper is to define maximal temperature and maximal power dissipation, which can be dissipated by this heat sink. Power dissipation depends at switching frequency. We can use results, which we obtained from simulations, for construction of high power frequency converter.

Keywords: heat sink, thermal field, cooling channel.

1 Introduction

In the frequency converter the power dissipation is arising. This power dissipation is converted into thermal energy. Energy must be fully spread out into a predefined heat sink. Therefore, we produced series of simulations of thermal fields to select the best construction solutions for heat sink KL 494.

3D model of the heat sink is created in 3D CAD program for this purpose. This model is equipped with power semiconductor modules.

Heat sink KL 494 is designed especially for the frequency converter in a locomotive. Its length is 630mm. Rated current to the power module is 100A and 280A and for short time it may be up to 560A. The type of power semiconductor module is SEMiX854GB176HDs and its dimensions are 57x155mm.

Mathematical expressions of all functions and precise calculations can be found in the literature [1-5]. By using these mathematical functions we compiled model. This model is shown in the following figure.

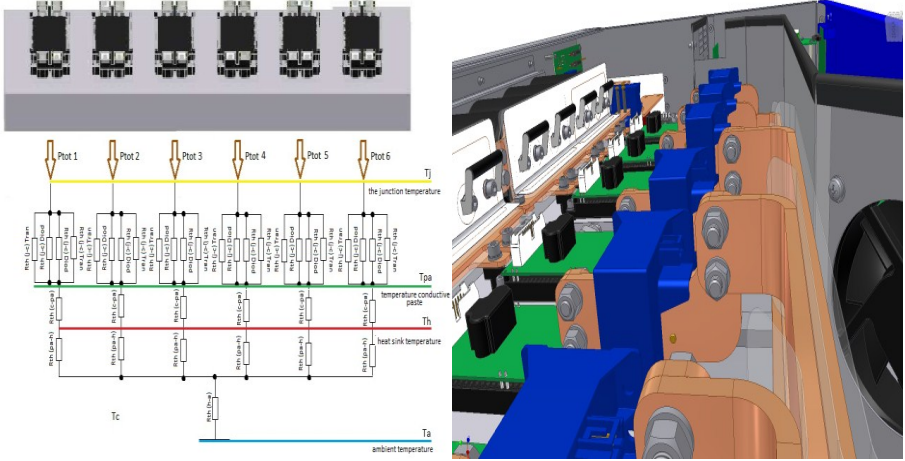


Fig. 1: Alternative scheme of convection thermal energy transfer and detail of construction of the converter [6]

2 Simulation results

Simulations are divided by power dissipation and number of cooling channels. The power modules are not considered in the simulations, because we cannot determine its inner construction.

Simulations are created in multi-physic program with function CFX and Steady State. The initial temperature of cooling liquid was set to $40^{\circ}C$. The overall flow rate was always set to $10L \cdot s^{-1}$. So if there are more cooling channels in the heat sink, flow rate in one channel is equal share of overall flow rate.

First simulation is created with two cooling channels in the heat sink of type KL 494. One channel is inlet and second channel is outlet. Cooling medium is water and its temperature is set to $40^{\circ}C$. We can see the results of these simulations in figure 2, 3 and 4. Dissipation power is lower on the left than on the right. All results we can find in Table 1.

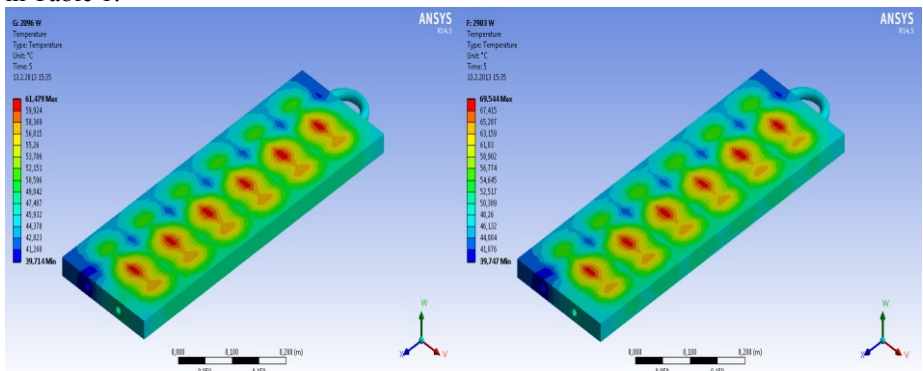


Fig. 2: Thermal field of heat sink with power dissipation 2039 and 2983 W .

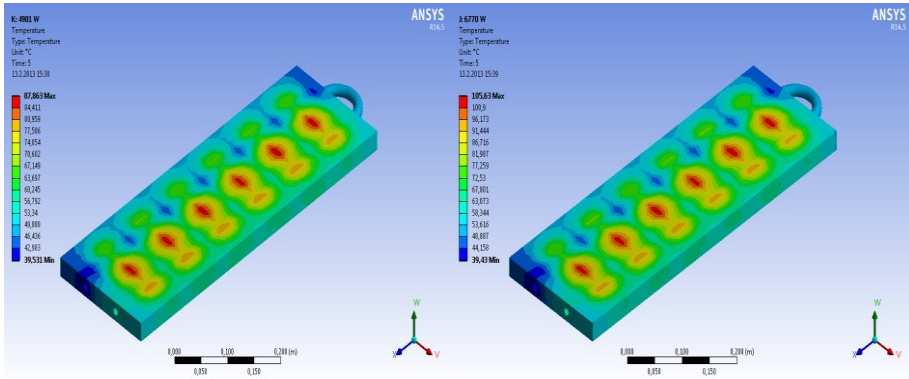


Fig. 3: Thermal field of heat sink with power dissipation 4091 and 6770 W.

Table 1: Table of maximal temperature of heat sink and temperature of cooling medium

Ptot [W]	Iin [A]	F [kHz]	Ta [C]	Tmax [C]	Tkap [C]
2039	100	1	40	61,479	48,126
2983	100	3	40	69,544	51,871
4901	100	5	40	87,863	59,663
6770	100	7	40	105,63	67,917

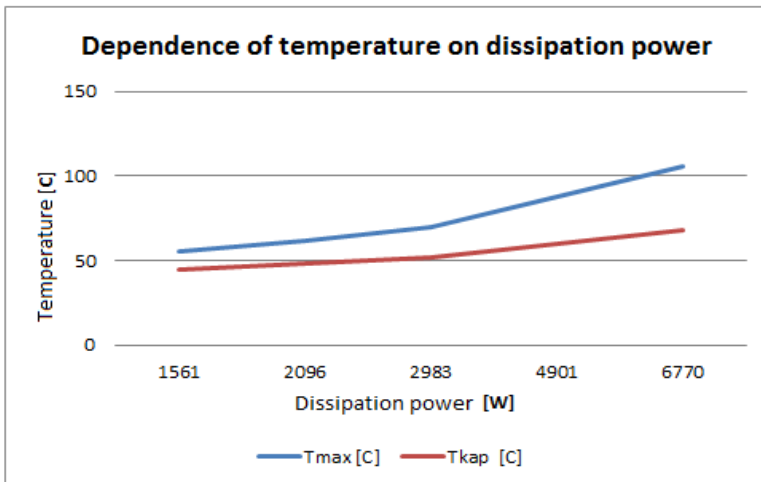


Fig. 4: Dependence of temperature on power dissipation

Next simulation was performed on 3D model of heat sink with four cooling channels. Heat sink is divided in two similar parts. In these parts we used two cooling channels. The first channel was inlet and the second was outlet. Flow rate was set on

$5l.s^{-1}$ per inlet channel. The temperature of cooling medium was set to $40^{\circ}C$. We can see the results of these simulations in figure 5, 6 and 7. Dissipation power is lower on the left than on the right. All results we can find in Table 2.

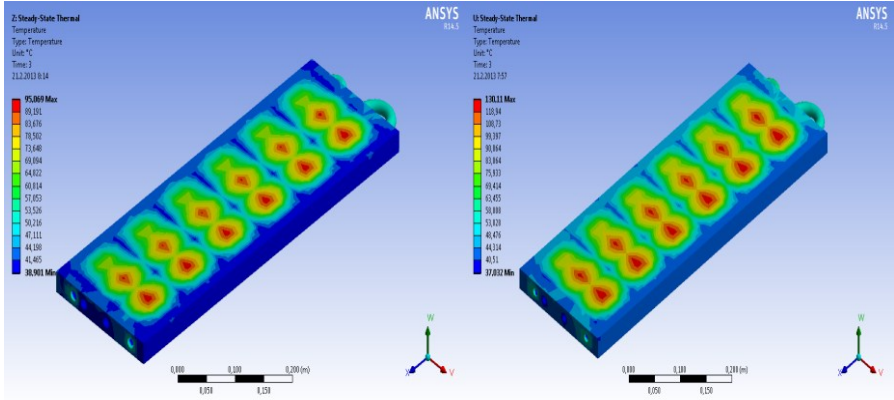


Fig. 5: Thermal field of heat sink with power dissipation 7490 and 9307W.

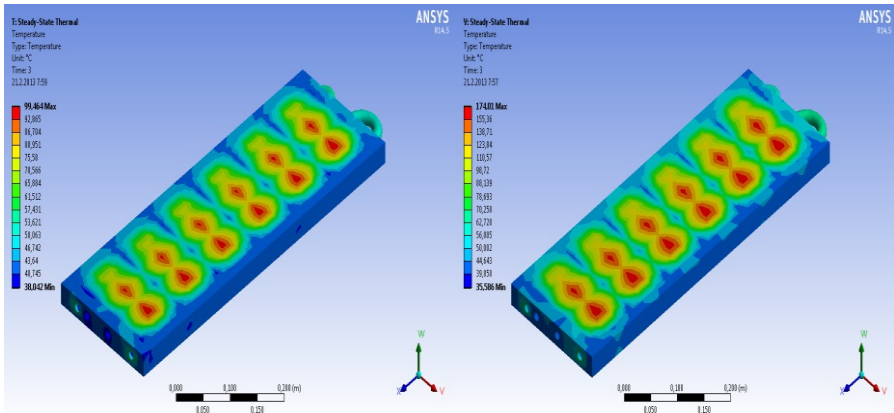


Fig. 6: Thermal field of heat sink with power dissipation 14110 and 20986W.

Table 2: Table of maximal temperature of heat sink and temperature of cooling medium

Ptot [W]	Iin [A]	F [kHz]	Ta [C]	Tmax [C]	Tkap [C]
7490	560	1	40	99,464	58,113
9307	560	3	40	99,464	58,113
14110	560	5	40	130,11	65,569
20986	560	7	40	174,01	78,891

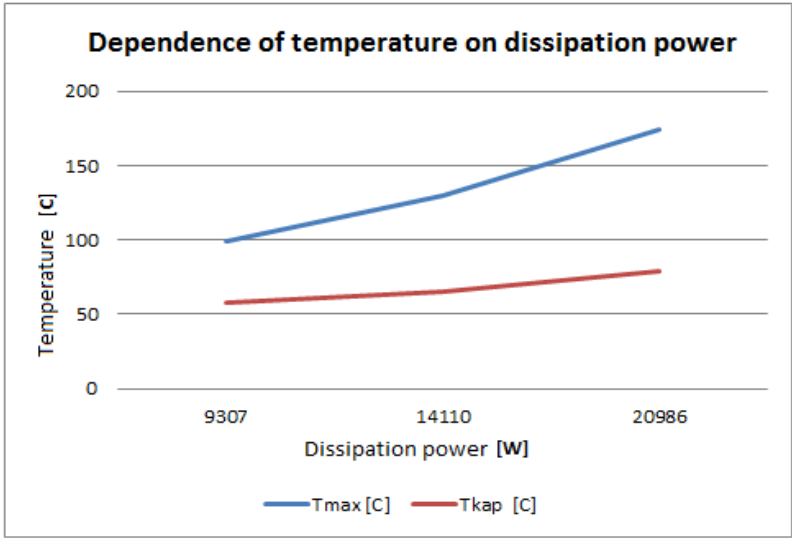


Fig. 7: Dependence of temperature on power dissipation

Last simulations have a new prototype cooling. Two pieces of aluminum blocks are mounted on the heat sink. Those blocks are used for distribution of liquid. The flow rate of cooling medium is divided into two halves. Inlet is represented by two independent cooling channels in middle and outlet is represented by two channels on sides. The flow rate of cooling medium is set to $10l.s^{-1}$. We can see the results of these simulations in figure 8. Dissipation power is lower on the left than on the right. All results we can find in Table 3.

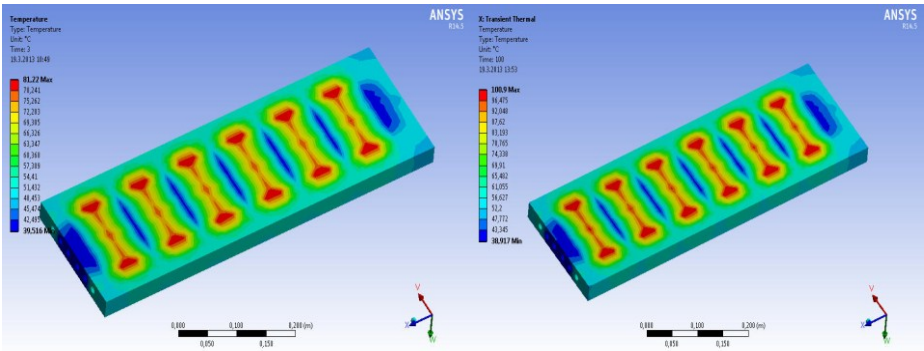


Fig. 8: Thermal field of heat sink with power dissipation 9307 and 20986W.

Table 3: Table of maximal temperature of heat sink and temperature of medium

Ptot [W]	Iin [A]	F [kHz]	Ta [C]	Tmax [C]	Tkap [C]
9307	560	3	40	81,22	53,189
20986	560	5	40	106,9	71,106

3 Conclusion

Simulations, which we created, give us an overview about the maximum temperature in the liquid cooled heat sink KL 494. All results are listed in Tables 1, 2 and 3.

In the first simulation, maximal power dissipation is set to $6770W$ and maximal temperature was $105,63^{\circ}C$. Flow rate was set to $10l.s^{-1}$ and we used only two cooling channels for this simulation. In the next simulation maximal power dissipation was set to $20986W$ and we used four cooling channels. Overall flow rate was set to $10l.s^{-1}$ per inlet cooling channels ($5 l.s^{-1}$ each channel).

From these simulations we created next type of cooling system. We added two block of aluminum to the heat sink. These blocks are installed on both sides of the heat sink. These blocks are used to distribute and modify the flow of cooling medium in the heat sink KL 494. Maximal temperature was $106,9^{\circ}C$ with power dissipation $20986W$.

Acknowledgement

The research described in this paper was supported by these projects:

- SGS project SP2013/80 – Utilizing of Modern SiC Components for MF – Power Semiconductor Converters

References

1. Chlebis, P., Havel, A., Vaculik, P., Pfof, Z.: Modern Instruments for increasing the Efficiency of the Energy Transfer in Electric Vehicles, 14th EPE-PEMC 2010 International Power Electronics and Motion Control Conference, September 6-8, Ohrid, Republic of Macedonia 2010, vol. 14., p. 89-93, ISBN 978-1-4244-7854-5
2. Brandstetter P., Chlebis P., Palacky P.: Direct Torque Control of Induction Motor with Direct Calculation of Voltage Vector. Advances in Electrical and Computer Engineering, vol. 10, n. 4, pp. 17-22, 2010, ISSN 1582-7445.
3. Chlebis, P., Dudek, J.: Comparison of Power Losses and Quality of Output Voltage from Different Types of Inverters. In 35th Annual IEEE Power Electronics Specialists Conference, Aachen, Germany, 2004, VOLS 1-6, pp. 3120-3126, ISBN 0-7803-8399-0.
4. Jaluria Y., Torrance K.E.: Computational Heat Transfer. Taylor and Francis, New York, 2003 ISBN 1-56932-477-5
5. Patočka, M., Matematická analýza signálů sinusové PWM, Sborník XXIV. Celostátní konference o el. pohonech, Plzeň 1995, str. 233-238.
6. Vaculík P., Havel A., Tvrdón M., Hromjak M., The Design and Construction of a Reversible Voltage Inverter Prototype. EPE 2012. Brno 2012, ISBN : , 978-80-214-4514-7.

Analog Electronic Circuits with Functional Blocks Based on Current Conveyors

Radek Šulák

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
radek.sulak@vsb.cz

Abstract. The main aim of this paper is simulation of several electronic circuits, in which new functional structures are used instead of traditional voltage operational amplifiers. These new functional structures contain of the only commercially available current conveyor, which is the non – invert positive current II. generation conveyor (CCII+). These new functional structures are based on the essential idea that “every active element can be described by a set of controlled sources and that the II. generation current conveyor is an universal constructional active element. In this paper, the current conveyors are firstly theoretically described and then used for construction of new functional structures. These new functional structures are consequently demonstrated in several simulations. As an example, a simulation of an active filter and Wien oscillator are presented in this study.

Keywords. current conveyor, wien oscillator, active filter

1 Introduction

In 1966, the concept of an active element was introduced, which surpassed the characteristics of operational amplifiers. This first conception is called I. generation of current conveyors. A few years later, II. generation of this element was introduced. Nowadays, it can be only found as part of another complex structure. If this element is implemented in electronic circuits, many advantages can be found, such as: it significantly improves dynamics of the circuit (increases slew rate), it increases frequency, at which circuit can operate (up to several hundred MHz), it extends frequency range of processed signals, it improves circuit's noise immunity at low supply voltage, and most importantly, it can operate at very low supply voltage, which can be very beneficial for circuits, which are powered by batteries. This article is based on the essential idea that every active element can be described by a set of controlled sources and that the II. generation of current conveyor is an universal constructional active element.

2 Current conveyors

Schematic symbol of the current conveyor is shown in Fig.1. Most commonly, the conveyors have only three terminals. These terminals differentiate from each other by function and properties [1]. The terminal Y is characterized as a high – impedance non-inverting voltage input. From this terminal, the voltage is conveyed to the terminal X (current input) by voltage follower but here it can also be considered as the output voltage terminal. However, if we consider that terminal X is input current terminal, the current is conveyed to the terminal Z by the current follower. The terminal Z is current output terminal of current conveyor [2].

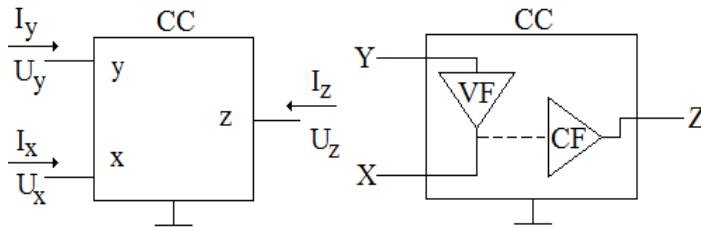


Fig. 1 Schematic symbol and block scheme of current conveyor

The relationship between input and output terminals [1]:

$$u_x = a \cdot u_y, i_y = b \cdot i_x, i_z = c \cdot i_x \quad (1)$$

Every current conveyor can be described by a suitable choice of the transmission coefficients of currents and voltages (a , b , c).

The coefficient „ a “ specifies if the current conveyor is inverting or non – inverting. If coefficient $a = 1$, the voltage that is conveyed from the terminal Y to the terminal X is not inverted and marking of the conveyor is CC. If coefficient $a = -1$, the voltage that is conveyed from the terminal Y to the terminal X is inverted and marking of conveyor is ICC [2].

The next coefficient „ b “ specifies generation of current conveyors. If coefficient $b = 1$, the current I_x is conveyed to the terminal Y with positive sign. Marking of conveyors is CCI. If coefficient $b = 0$, the current I_x is not conveyed to the terminal Y and marking of conveyors is CCII. If coefficient $b = -1$, the current I_x is conveyed to the terminal Y with negative sign and marking of conveyors is CCIII [2].

The coefficient „ c “ specifies behavior of the current conveyors. Also, terminology of the non – inverting (positive) and inverting (negative) current conveyor corresponds to the orientation of the output current. Hence, when the coefficient $c = 1$, it is a positive current conveyor and it is marked with a plus sign (CCII+). When the coefficient $c = -1$ it is a negative current conveyor and it is marked with a minus sign (CCII-) [2].

Another important element is the universal current conveyor (UCC), which has been developed on the basic idea that all the previously mentioned types of current conveyors can be implemented into only one structural element.

3 Basic structure with current conveyors

Every active element can be generally described with ease, regardless on what is the structure or what are the components of the required circuit. The best description is by a set of controlled sources, which can describe all the structures [3].

Furthermore, due to the fact that the non – inverting positive conveyor of II. generation (CCII+) is the only commercially available current conveyor, its use can be also assumed as of the universal active element. However, this applies under certain conditions.

Therefore, based on the claim that CCII+ is an universal building active element [3], it is practically possible to realize four basic functional block structures. These are: a voltage source controlled by voltage, a current source controlled by voltage, a voltage source controlled by current, a current source controlled by current.

The structure, which represents compensation for operational amplifier is shown in Fig.2 [3]. The voltage terminal Y1 represents the inverting input of this structure. The voltage input Y2 represents the non – inverting input of this structure. The terminal Z2 is output of this structure. Unused terminals are grounded. The output is completed with a voltage follower [4].

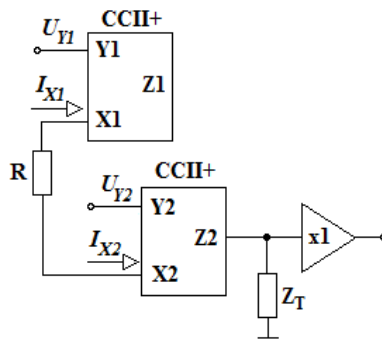


Fig. 2 Functional block structure with current conveyors

The impedance Z_T and the resistor R are the most important elements in this structure. They significantly affect the structure and the proposed circuit as they determine the gain of the structure. The impedance Z_T ideally approaches to infinity and the resistor R ideally approaches to zero.

$$A_s = \frac{Z_T}{R} \tag{2}$$

4 Application of functional structures in electronic circuits

In this paragraph, the above mentioned structure (Fig.3) providing the replacement of an operational amplifier by current conveyors is used in a design and construction of low – pass filter of 6th order in a non – inverting connection. After removing of the

operational amplifier, the structure is directly fitted on the place of the original operational amplifier and that without any further modifications of the circuit.

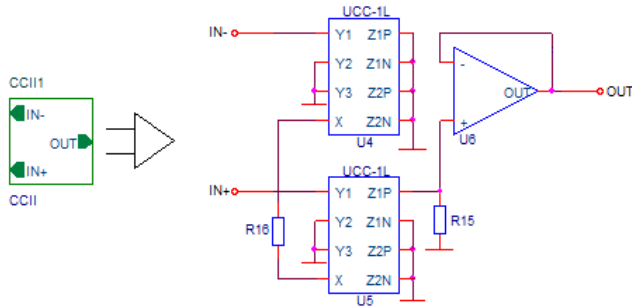


Fig. 3 The compensation for operational amplifier

The structure is composed of two universal current conveyors, which are connected as a positive current conveyor of II. generation. It is placed instead of operational amplifier. On this sample, we can confirm the assertion that with the universal current conveyors we can realize a current conveyor of any generation. The proper function of the low-pass filter can be then confirmed by provided graphic waveforms (Fig.3). The voltage inputs of universal current conveyors represents inverting and non – inverting inputs. The current output of universal current conveyor U5 is output of this structure which is connected to the input of voltage follower. The resistor R15 is connected at the input of this follower, which together with resistor R16 determine the gain of the structure. Unused terminals must be grounded.

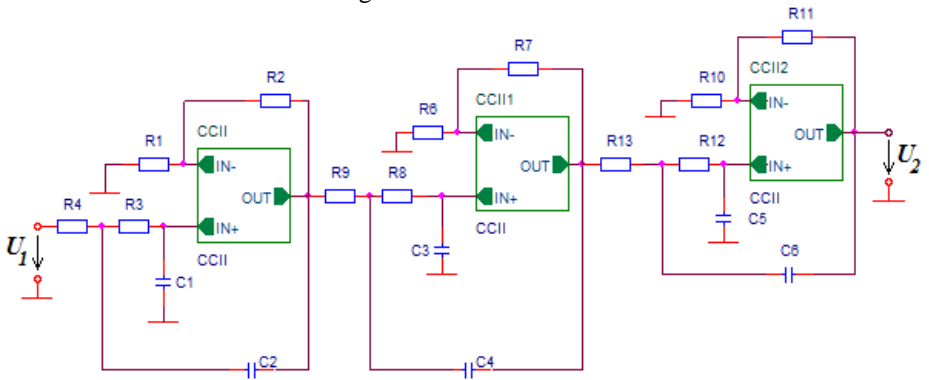


Fig. 4 Connection diagram of low – pass filter of 6th order with current conveyors

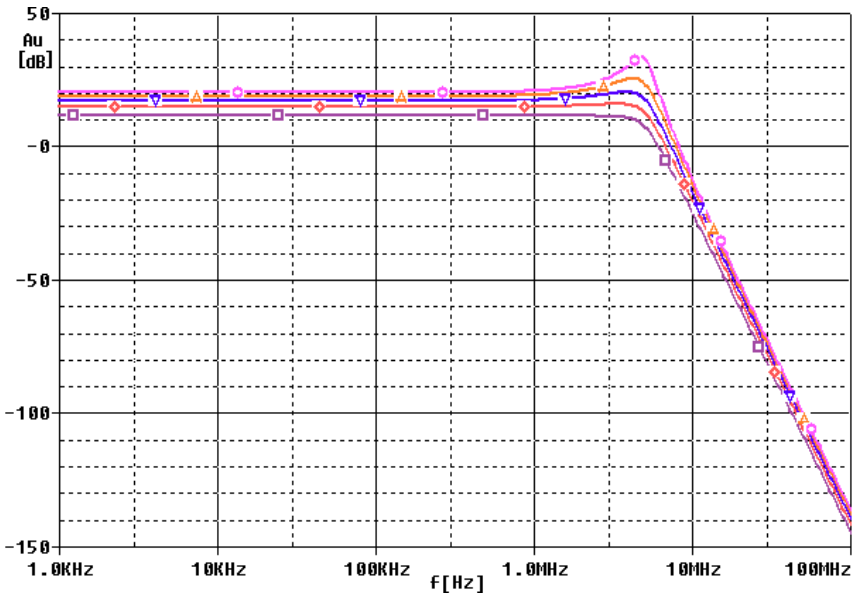


Fig. 5 Amplitude frequency characteristic of low – pass filter of 6th order ($f_c=5\text{MHz}$)

Furthermore, the functional block structure is used in Wien’s oscillator [2] (Fig.3). Also, in this case, the classic operational amplifier is replaced by the current conveyor. The oscillating frequency was chosen to $f_G = 1\text{MHz}$.

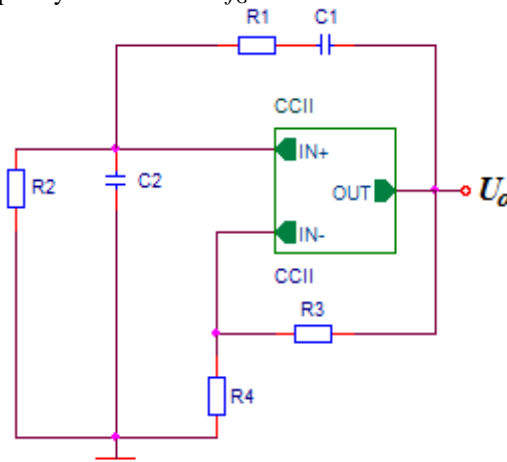


Fig. 6 Wien oscillator

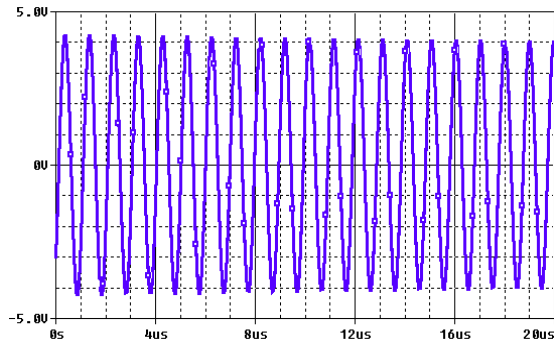


Fig. 7 Waveform of Wien oscillator ($f_G = 1\text{MHz}$) with replaced structure

5 Conclusion

The paper describes current conveyors and their using in electronic circuits. It is confirmed that the current conveyor of II. generation can operate on very high frequencies and can be used in the construction of new structures as a universal building block. This can be also seen in the provided waveform graph. Based on these results, the current conveyors have a high potential to be a strong competitor to the traditional operational amplifiers. Especially in case of frequency dependent circuits.

Acknowledgement

In the paper there are the results of the project SP2013/118 which was supported by Student Grant Competition of VSB - Technical University of Ostrava. This research was supervised by Prof. Ing. Pavel Brandštetter, CSc.

References

1. ČAJKA, Josef, Tomáš DOSTÁL a Kamil VRBA. *New terminology and unifying view of the current conveyors* [online] (in Czech language). 2001[cit. 2012-04-30]. Available from: <http://www.elektrorevue.cz/clanky/01024/index.html>
2. KLEIN, Lukáš. *The current and voltage conveyors*. Additional instructional text,(in Czech language) Ostrava: Vysoká škola báňská – Technická univerzita, 2011.
3. PUNČOCHÁŘ, Josef. *History and present of operational amplifiers*. BEN. Praha, 2002. ISBN 80-7300-047-4.
4. BRANDŠTETTER, P. KLEIN, L.: *Applications of Non-Inverting Positive Second Generation Current Conveyor as a Commercially Available Versatile Active Element*. Conference Proceedings of International Conference on Signals and Electronic Systems - ICSES'10. Gliwice, 2010. pp. 157-160. ISBN 978-1-4244-5307-8.

Simulation of the Three Phase DC/DC Converter used for Designing and Testing of Traction Battery Models

Martin Tvrdon and Ales Havel

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{martin.tvrdon, ales.havel}@vsb.cz

Abstract. This paper provides an introduction to the current work, which was done on the developing of battery models used as a main energy source in traction for the propulsion of electric vehicles. The first chapter presents a currently built prototype of the charging station, which will use various battery models embedded in its control and regulation structure. The simulation program for the design and testing of these models is presented in the next chapter of this article. The paper concludes with an application schema which is formed in the prototypes of charging stations.

1 Introduction

The power system in world's most advanced agglomerations is undergoing a major transformation. Today's power system transforms itself into the power system of the future with a focus on a high proportion of electricity production from renewable energy sources (RES), the accumulation of electrical energy in a dense network of various storage units with various capacities and also for the production of electrical energy at the point of consumption (e.g. CHP units – power sources using several types of energy – electricity or thermal energy used for residential spaces heating). The points listed above are associated with the production and distribution of electrical energy. The following points provide time-spread power consumption and do not generate peak consumption in the power system. For example, appliances with accumulative character, or appliances which do not need immediate operation are switched according to the requirements of the power grid.

New types of charging stations for electric vehicles (especially widespread in the future) with high power requirements can be divided into two main groups according to the energy amount; the immediate fast charging or slow charging, when the consumer chooses the charging time after which the vehicle will be ready for operation. The second method allows more efficient utilization of power grid capacity for immediate and fast charging requirements. [1], [2]

Another effort in building power grids of the future is to prevent widespread electricity outages, so-called "blackouts". These failures will be reduced by the structure of the grids, which will be more stable without large fluctuations in electricity consumption, etc. Another level of providing supply of electricity will be the possibility to

operate in island modes of certain sections of power grid, where the power sources will be made by accumulation units or even by the EV batteries connected to the charging stations. These grids of the future were given the general name of Smart Grids - Intelligent networks.

This paper deals with the accumulation of electrical energy in storage units of small and medium power and with the direct transfer of electricity with high instantaneous power via fast charging stations into electric vehicles. The current fast charging stations require high power connection to the grid and its use leads to large peak demands. This adverse effect on the electrical grid is eliminated by the concept of drawing energy directly from the accumulation units. The unit itself also creates a control point of the energy consumption by accumulating the energy without burdening the power grid by the unnecessary supply at surplus times, even with the addition of renewable energy source with time-inefficient operation. The coupler between the accumulation unit and an electric vehicle is formed by the three-phase DC/DC converter.

2 The Prototype of Fast Charging Station

The prototype of fast charging station consists of a three-phase DC/DC converter, which controls the electrical parameters of the values in order to satisfy proper battery charging of electric vehicles.

The multiple-branch structure of the converter was chosen because of the distribution of transferred power into multiple branches, which relieves the performance of individual IGBT modules. Other advantages of this topology are in the distribution of power dissipation on greater cooling surface allowing easier and faster cooling of the converter, reduce of the voltage ripple on the capacitor battery and also reduce of the load current ripple.[1], [4]

From a theoretical point of view, this converter is considered to be built from three separate parallel connected buck converters with commonly overlapped control. From this perspective it is necessary to define basic control terms of such converter. The output voltage is dependent on the duty cycle or relative switch which is commonly labeled as z . The output voltage of the converter in the case of continuous current is then given by:

$$U_{OUT} = z \cdot U_{IN} \quad (1)$$

Ideally, when the parameters of components in each of the parallel branches are the same and all branches of the converter are controlled with the same duty cycle z , then the output voltages of each phase are with the respect to time shifts the same, respectively, the average voltage values of each phase are equal.

With the respect to the real parameters of the circuit and to the manufacturing tolerances of individual components, the currents in each phase of the circuit may differ. In order to ensure uniform current distribution on the branches of the converter, it is necessary to provide current regulation in individual branches separately. This will also cause uniform distribution of the power. This control is ensured by the control system of the converter.

The converter structure of this type requires a special way of control. Output voltage of the converter is adjusted by changing the duty cycle of the transistor known as z , which was specified in the equation (1). The voltage of each branch is kept on the constant level, so the current in all branches will be the same. The regulation is provided in each branch separately to the desired value of the output current $I_{OUT}/3$. This way of control ensures even spreading of the transferred power to the individual branches. In order to realize additional benefits of this structure, it is necessary to divide switching times of the transistors in individual branches. This means to create a commonly overlapped control of the converter, which is possible to produce by a constant switching frequency control, is in this case generated by three saw wave signals with mutual shift of 120° .

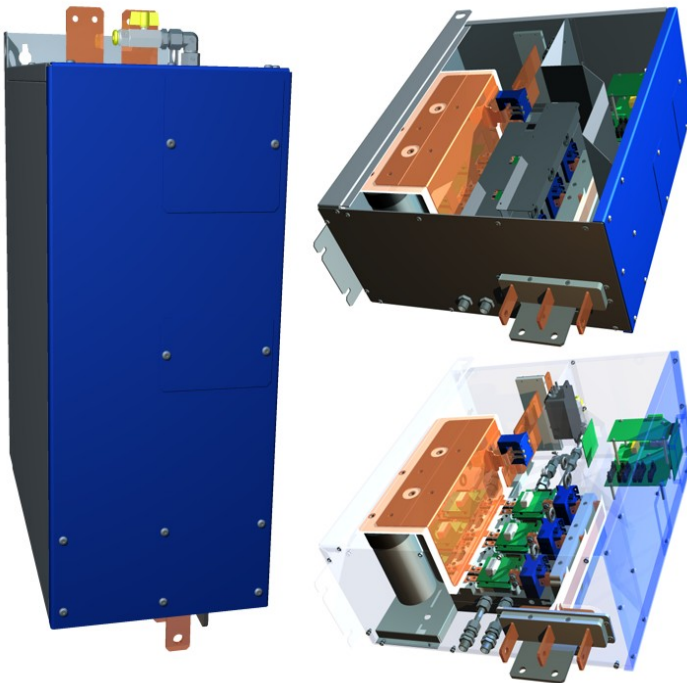


Fig. 1. The prototype of power converter used in fast charging station for electric vehicles [3]

3 Simulation of the Three Phase DC/DC Converter

The simulation schematic below tries to closely describe the real model of fast charging station. It allows the monitoring of all circuit parameters in different parts of the power circuit. The model of the battery from Simscape library in the examined part of the power circuit is shown in Fig. 2. Changing this model allows to develop and test different models of batteries. Developed models can be implemented in the control system of fast charging station. The reason of real battery model implementation is given in the following chapter.

The simulation schematic is composed of parts from the Simscape library. These parts are representing real electrical equipment. The system is powered by primary battery of the type LiFePO_4 . The next part represents a capacitor bank, which supplies power for dynamic states caused by switching of the converter. The following part represents the three-phase DC/DC converter, which is composed from IGBT switching modules and its each phase is controlled separately. Converter's outputs are connected to charged battery through three reactors. Charged battery is also of the type LiFePO_4 . The main simulation system is equipped with sensors to record the values for further evaluation of the circuit.

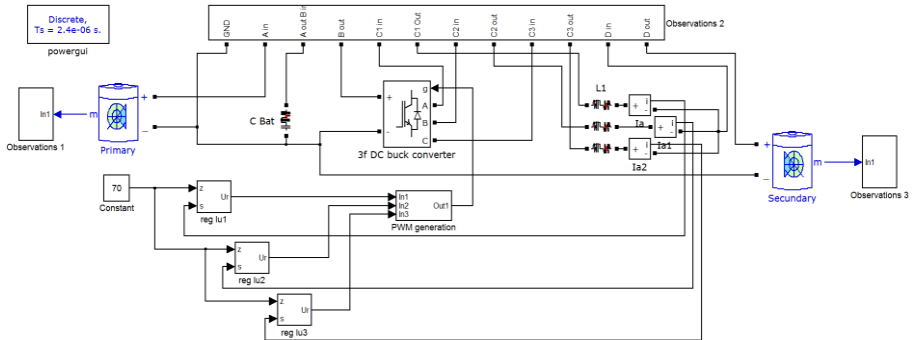


Fig. 2. Simulation scheme of the DC/DC converter for designing and testing of battery models

4 Structure of the control system of fast charging station for EV

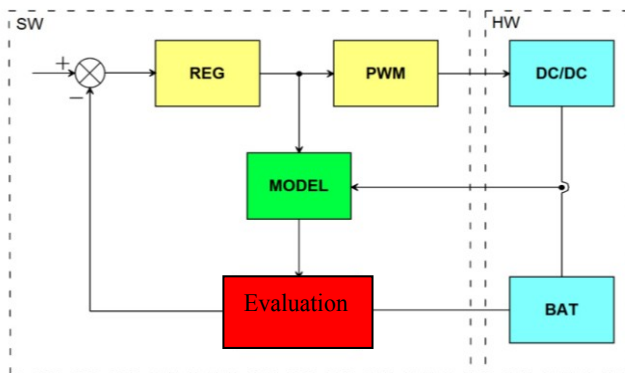


Fig. 3. Block diagram of the control system of fast charging station with the implementation of battery model to predict the behavior of charged batteries.

The structure of the control system for fast charging station is shown in Fig. 3. The scheme is divided into two main parts labeled HW and SW. HW section contains a three-phase DC/DC converter and a charged battery with necessary sensors. SW part is included in the control system and contains blocks called battery model and evaluation. These blocks represent the main focus of this research. The battery model block

provides information about inner battery charging status. These data are difficult or impossible to measure directly during the charging process. The evaluation block then evaluates and compares the status of the model and batteries. The decision of following charging parameters is then based on the results of the comparison. The necessary regulation and production of pulses for switched converter is ensured by blocks labeled REG and PWM.

Conclusion

This work follows a more focused development of battery models and control algorithms for fast charging stations with a focus on the implementation of battery models into control structures of the converter. This area of research is especially promising because of the acceleration of charging, increasing of the charging efficiency and extension of the life of charged batteries. Designed and implemented solution is applicable in a wider structure of charging stations for electric and hybrid electric vehicles.

Acknowledgements

The research described in this paper was partially supported by these projects:

- SGS project SP2013/80 – Utilizing of Modern SiC Components for MF – Power Semiconductor Converters.
- ENET CZ.1.05/2.1.00/03.0069 – Energy Units for Utilization of non Traditional Energy Sources.

References

- [1] CHLEBIS, Petr and Jan DUDEK. Comparison of Power Losses and Quality of Output Voltage from Different Types of Inverters. In *35th Annual IEEE Power Electronics Specialists Conference*, Aachen, Germany, 2004, VOLS 1-6, p. 3120-3126, ISBN 0-7803-8399-0.
- [2] JUNG-SONG, Moon, Lee JUNG-HIO, He IN-YONG, Lee TEACK-KIE and Won CHUNG-YUEN. An Efficient Battery Charging Algorithm Based on State-of-charge Estimation for Electric Vehicle, *Electrical Machines and Systems (ICEMS), International Conference on*, August 2011, Beijing, China, p 1-6, ISBN: 978-1-4577-1044-5.
- [3] VACULIK, Petr, Ales HAVEL, Martin TVRDON and Michal HROMJAK. The Design and Construction of a Reversible Voltage Inverter Prototype. *13th International Scientific Conference Electric Power Engineering 2012*, May 23 – 25, Brno, Czech Republic, 2012, p. 1059-1062. ISBN 978-80-214-4514-7.
- [4] CHLEBIS, Petr, Ales HAVEL, Petr VACULIK and Zdenek PFOF. Modern Instruments for Increasing the Efficiency of the Energy Transfer in Electric Vehicles, *14th EPE-PEMC 2010 International Power Electronics and Motion Control Conference*, September 6 – 8, Ohrid, Republic of Macedonia, 2010, vol. 14., p. 89-93, ISBN 978-1-4244-7854-5.

Compression and Multiple Range Query in Tree Data Structures

Václav Bašniar

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
vaclav.basniar@vsb.cz

Abstract. In this paper we introduce an application of the fast decoding algorithms for variable-length codes - Fibonacci of order 2 and 3, Elias-Fibonacci, and Elias-delta - and multiple range query algorithm for efficient querying of the B-tree data structure with a vector of queries. The application of both methods at once is not a subject of this paper. If we have a large vector of queries, for example a vector of n-grams extracted from a document, usage of a simple point or range query algorithm isn't very effective because of their redundant accesses to main memory or secondary storage. The multi range query resolves this problem.

Compression methods can reduce all costs associated with memory space and data transfer. In the case of indexing data structures the compression/decompression time is crucial due to many accesses to values stored in a data structure. Therefore, lightweight compression algorithms have been applied in this area. However, the lightweight compression algorithms are not always appropriate for data structure compression. Although universal variable-length codes are an important class of codes, they have not been yet applied in the area of data structure compression.

Keywords: variable-length coding, querying, tree data structures, B-tree, R-tree, compression

1 Introduction

When the tree data structure is queried with the vector of queries the simple point and range query algorithms process each query one by one separately and thus they invoke some nodes from cache multiple times generating a lot of overhead. In Section 2 is presented multiple range algorithm for B-Tree data structure that invokes each node only once during a whole querying process.

In Section 3 is presented compression of B-tree and R-tree data structures using fast decoding algorithms. As stated above the compression efficiency for tree data structures is highly dependant on the compression/decompression time due to many accesses to stored values and that the lightweight compression algorithms, e.g. RLE, LZ, dictionary methods, Huffman coding etc. [6], have been applied in this area [4, 3]. These methods aren't always appropriate for data structure compression. For example, RLE is applicable only for data with

many repeating values or it is not possible to construct Huffman codes for a large number domain, e.g. 32 bit-length, since the Huffman tree becomes large [6], moreover, the bit-length of codewords is driven by frequencies of input numbers which limits performance of update operations, and so on.

The universal variable-length codes have not been yet applied in the area of data structure compression, since the overhead of encoding/decoding algorithms is rather high. We have applied fast decoding algorithms for Fibonacci of order 2 and 3, Elias-Fibonacci, and Elias-delta that have been introduced in [7, 5].

2 Multiple Range Query Algorithm

Main idea of the multiple range query presented in [2] is to minimize count of costly accesses to the tree data structures' nodes while processing large vector of point and range queries. This minimalization leads to reduction of the memory hit count and more importantly to reduction of the memory miss count.

A multiple range query algorithm is capable of accessing each node only once. This is achieved by comparison of relevant subsets of the query vector to the items stored in each node. The algorithm uses breadth-first method to proceed through the tree data structure. In each node the algorithm does the comparison and constructs new subset of relevant query vector for each of its child and recursively. This is the procedure for inner nodes of the tree data structure. For leaf nodes the algorithm simply compares query vector with the items in the node and returns those which are relevant. This algorithm can be used for B-tree and R-tree data structures.

But for the B-tree data structure the algorithm can be slightly altered. As the B-tree preserves order of stored items we don't have to compare each item with each query like in the R-tree. Instead we can use a bisection method to find minimal item for a query and then sequentially determine which queries are relevant to the child node. This can significantly reduce the complexity.

Algorithm 1 describes an implementation of the inner node's processing part of the Multiple range query algorithm. As we can see it uses two lists for relevant queries. One contains queries relevant for the child node - $nrqs$ - and one with the queries for the current node - rqs . Algorithm removes no longer relevant queries from rqs and adds those which are relevant to the child node to $nrqs$. When it hits the last query relevant to a child node it returns to the adjacent method, where the recursive scan through the B-tree happens.

3 Data Structure Compression Using Variable-Length Codes

If we want for example to index a large collection of data using a method that produces even larger datasets we can easily run out of main memory space. In the case when the operation system starts to use a secondary storage the efficiency

Algorithm 1: Batch ScanNode Algorithm – B-Tree

```

Input : out startOrder – the order of the item of the node to be scanned, rqs – an array of
range queries
Output: nrqs – an array of all range queries relevant for the first relevant subtree found

1 nrqs.Init();
2 while startOrder < itemCount do
3   proceed ← false;
4   startOrder ← FindMinimalOrder(startOrder, rqs [0]);
5   for i ← 0 to rqs.Count() do
6     isIntersectedResult ← IsIntersected(GetItem(startOrder-1), GetItem(startOrder),
rqs [i]);
7     if isIntersectedResult = 0 then
8       nrqs.Add(rqs [i]);
9       proceed ← true;
10    end
11    else if isIntersectedResult < 0 then
12      rqs.Remove (rqs [i]);
13    end
14    else
15      proceed ← true;
16      break;
17    end
18  end
19  if proceed then
20    startOrder ← startOrder +1;
21    break;
22  end
23 end

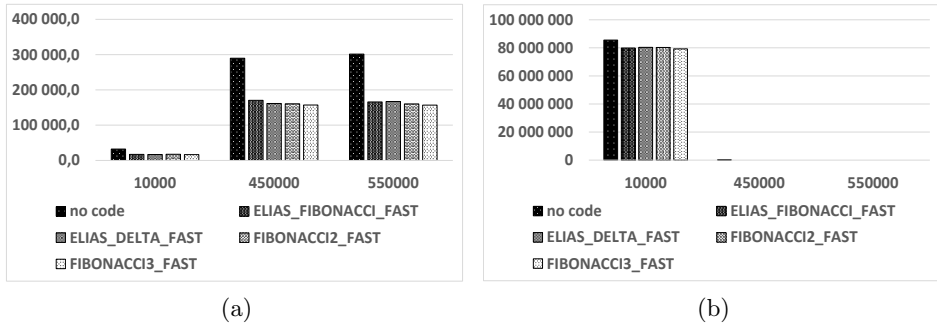
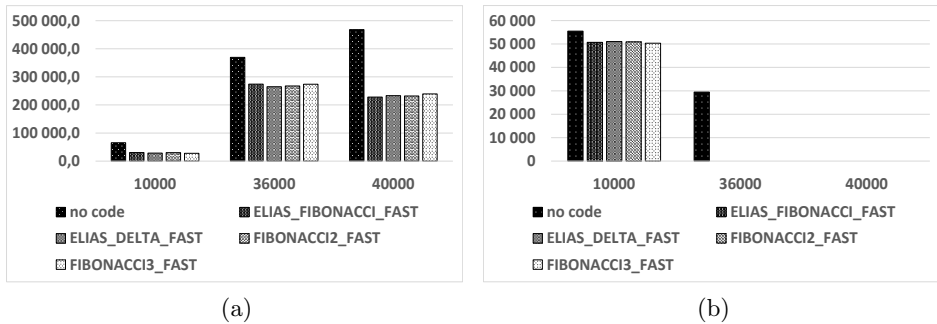
```

of the data structure used for indexation will start to degrade. As a solution we tried to use a data compression.

The variable-length codes Fibonacci of order 2 and 3, Elias-Fibonacci, and Elias-delta were used for compression. Since 99% to 99.9% of all nodes in B-tree data structures are leaf nodes [4] and that the shape of data from pair (key, data) vary with the usage of B-tree we have decided to code only keys in leaf nodes. However, this method can be applied also on R-tree data structure.

In the following graphs (Fig. 1 through Fig. 4) we can see statistics from testing of variable-length codes application on B-tree and R-tree data structures. As a data we have used iDoc2011 collection that contains thesis defended in VŠB - Technical University of Ostrava in 2009 through 2011 and Words collections that represents so called bag-of-words. The iDoc2011 was used to extract 35,952,666 tuples of dimension 4 which were then inserted in B-tree. As a query was used randomly chosen document from the collection. The document was used for creation of range query vector which retrieve average result of size of 1.5 items. The Words contains 483,450,157 tuples of dimension 3 these were inserted in R-tree data structure. As a query was used a set of randomly chosen tuples from the Words collection.

Tests were done using different sizes of node cache so a case when all of the used methods accesses the secondary storage, a case when no method uses a

Fig. 1. iDoc2011 Collection Statistics - (a) Inserts/s, (b) Disc Access Cost for Insert**Fig. 2.** iDoc2011 Collection Statistics - (a) Queries/s, (b) Disc Access Cost for Querying

secondary storage and a case in which only method without coding accesses the secondary storage can be demonstrated.

As we can see in most of cases the method with no coding was faster then methods with coding regardless of cache size with the exception in Figure 4(b). This can be explained so that the query vector for Words collection was heavily randomized from original collection's items thus the query algorithm needed to access a lot of different nodes. So that the DAC has in this case much more impact on the efficiency then the coding itself. In graphs can also be seen worse performance of methods in some cases when the cache is unnecessary big.

4 Conclusion

In this paper, we introduced the multi-range query algorithm and statistics for compression in tree data structures using variable length codes. The statistics for compression in tree data structures showed that the compression isn't very

Fig. 3. Words Collection Statistics - (a) Inserts/s, (b) Disc Access Cost for Insert

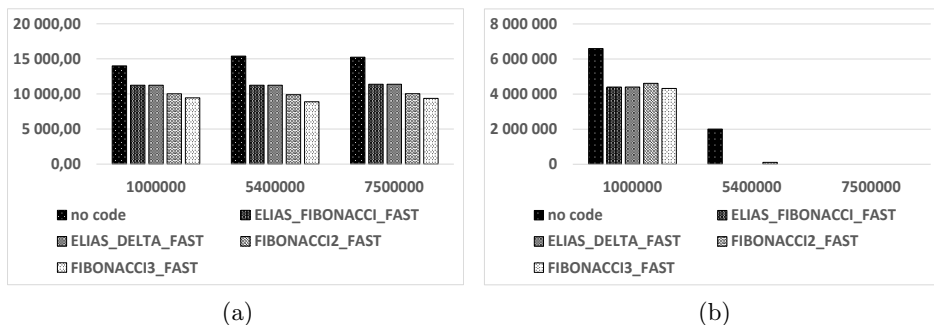
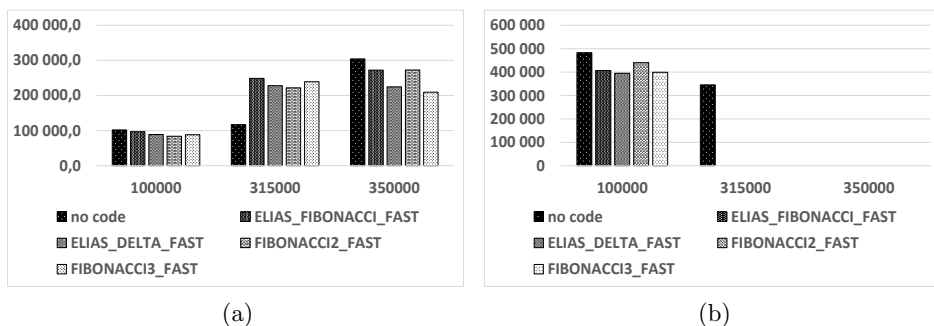


Fig. 4. Words Collection Statistics - (a) Point Queries/s, (b) Disc Access Cost for Point Queries



effective for the most cases. But if we need as much of the structure size in the main memory as possible the compression can be used.

5 Projects and Publications

Václav Bašniar is a member of team solving the grant of SGS No. SP2013/42, Detekce plagiovaných dokumentů (řešitel: Ing. Radim Bača, Ph.D.). He is co-author of one submitted journal article [1].

Own References

1. V. Bašniar and M. Krátký. Index-Based N-gram Extraction from Large Document Collections. *Submitted in Journal of Digital Information Management*, 2012.

Other References

2. P. Chovanec and M. Krátký. On the Efficiency of Multiple Range Query Processing in Multidimensional Data Structures. Submitted at IDEAS 2013.
3. D. Comer. The Ubiquitous B-Tree. *Computing Surveys*, 11(2):123137, 1979.
4. G. Graefe. *Modern B-Tree Techniques*. now Publishers Inc., 2010.
5. S. T. Klein and M. K. Ben-Nissan. On the Usefulness of Fibonacci Compression Codes. *The Computer Journal*, 53(6):701–716, 2010.
6. D. Salomon. *Data Compression The Complete Reference*. Fourth Edition, Springer-Verlag, New York, 2007.
7. J. Walder, M. Krátký, R. Bača, J. Platoš, and V. Snášel. Fast Decoding Algorithms for Variable-Lengths Codes. *Submitted in Information Science*, February, 2010.

Processing of Range Query in R-tree Using CPU and GPU

Pavel Bednář

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
pavel.bednar@vsb.cz

Abstract. Onedimensional or multidimensional range query is one of the most important query of physical implementation of DBMS. The number of compared items (of a data structure) can be enormous especially for lower selectivity of the range query. The number of compare operations increases for more complex items (or tuples) with the longer length, e.g. words stored in a R-tree. Due to the possibly high number of compare operations executed during the range query processing, we can take into account hardware devices providing a parallel task computation like CPU's SIMD or GPU. In this paper, we show different variant of the range query algorithm using CPU and GPU. These variants have different scalability and performance. We need to identify pros and cons of each algorithm and use advantage of each one in a DBMS kernel.

Keywords: range query processing, multidimensional range query, GPU, CUDA, rtree

1 Introduction

Range query (or range scan) [9] is one of the most important query of physical implementation of DBMS [7]. There are two basic variants: onedimensional and multidimensional range query. DBMS often utilize two types of data structures (and algorithms) supporting these range queries.

Onedimensional range query is often implemented in a data structure like a B-tree [2] and it can be processed in an execution plan of the following SQL statement: `SELECT * FROM T WHERE q1 ≤ T.attr1 ≤ qh1`. Multidimensional range query is often implemented by multidimensional data structures, e.g. n -dimensional B-tree [5], R-tree [6] or R*-tree [3] and it can be processed for the following SQL statement: `SELECT * FROM T WHERE q1 ≤ T.attr1 ≤ qh1 AND ... AND q1n ≤ T.attrn ≤ qhn`. In other words, this query retrieves all tuples of an n -dimensional space matched by an n -dimensional query rectangle.

In generally, there are no significant differences between one and multidimensional range queries; both range queries must compare individual values of a tuple. We aim our effort to the multidimensional range query in this work summarizing our article [1].

The R-tree is one of the most popular multidimensional data structure which is utilized in commercial database systems. This data structure bounds spatially

near points by multidimensional rectangles. It supports various types of queries, e.g. point and range queries. In the case of the R-tree, the range query returns tuples in a query multidimensional rectangle (QR).

A general structure of the R-tree is shown in Figure 1.

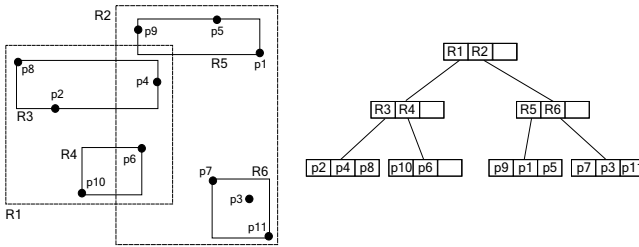


Fig. 1. Planar representation and general structure of the R-tree

It is a hierarchical data structure representing spatial data by the set of nested n -dimensional *minimum bounding rectangles* (MBR). Each MBR is defined by two tuples QL and QH , where $QL^i \leq QH^i$, $1 \leq i \leq n$. If \mathcal{N} is an inner node, it contains pairs (R_i, P_i) , where P_i is a pointer to a child of the node \mathcal{N} . If R is the inner node MBR, then the rectangles R_i corresponding to the children \mathcal{N}_i of \mathcal{N} are contained in R . Rectangles at the same tree level may overlap. If \mathcal{N} is a leaf node, it contains pairs (R_i, O_i) , so called *index records*, where R_i contains a spatial object O_i . Each node of the R-tree contains between m and M entries unless it is the root and corresponds to a disk page.

2 Multidimensional Range Query Processing

In the case of the multidimensional range query, we have two primitive operations: *IsInRectangle*, returning true if a tuple is in the query rectangle, and *IsIntersected*, returning true if a rectangle (or MBR – minimal bounding rectangle) intersects the query rectangle. These operations are used in both sequential as well R-tree range query implementations. Both these operations must perform two compare operations for each dimension. However, in [4], we show that it is not necessary to check all dimensions in the case of the conventional implementation and it is the reason why the SIMD implementation is not always successful.

2.1 Range Query Processing Using CPU

The n -dimensional range query returns all items of n -dimensional space (tuples or spatial objects) matched by the query rectangle (QR). The query rectangle is defined by two n -dimensional tuples QL and QH like in the case of the MBR,

and we require $QL \leq QH$. In the case of point data, a tuple is matched if it is in the QR. In the case of spatial objects in leaf nodes, an object is matched if its MBR is intersected by the QR.

The range query algorithm traverses the tree from the root node and follows only relevant items in each inner node. An item is relevant if its MBR is intersected by the query rectangle; the `IsIntersected` operation is invoked for this test. The range query algorithm search all subtrees recursively and it is finished after all relevant subtrees are processed. All matched items in leaf nodes are added to buffer. When buffer reaches its capacity all leaf nodes are tested. In the case of point data in leaf nodes, a tuple is added to the result if it is in the query rectangle (the `IsInRectangle` operation is used for this test). In the case of spatial objects in leaf nodes, an object is add in the result if its bounding rectangle is intersected by the query rectangle.

2.2 Range Query Processing using GPU

Our GPU range query algorithm is written in C++ and CUDA SDK for the compute capability 2.0 and higher [8]. In the area of GPU algorithms, a common technique is to arrange data for needs of the GPU algorithm. Since we suppose common row-oriented DBMS, input data are unchanged in the form of n-tuples. We do not consider another tuple arrangement, e.g. column-oriented [10].

In the case of the GPU algorithm we use similar technique for searching the R-tree. We use buffer as in case of CPU range query. For inner nodes we use same `IsIntersected` method as for CPU. At the moment we do not have available results for its GPU variant. The difference between CPU algorithm is that we need to transfer all leaf nodes in buffer to GPU's memory first. After all leaf nodes are transfered to GPU's memory we invoke GPU variant of `IsInRectangle` algorithm. The GPU algorithm cannot access CPU's variables so we cannot add corresponding items directly into result set. Instead we return only vector of informations which tuples should be added to result. Instead we fill result set later in CPU using `AddVectorToResultSet` method.

3 Experimental Results

In our experiments, we compare the performance of range query processing using CPU and GPU algorithms. In our test, we utilize POKER¹ collection. Characteristics of collection are shown in Table 1. We tested the performance for 40 range queries for each data collection divided to 4 query groups according to the selectivity. All range queries were 10× repeatedly executed and results have been averaged for one query. For searching the R-tree we used prefetching. We stored leaf nodes in buffer of capacity 512. When the capacity of buffer is reached we invoke searching on CPU or GPU.

¹ <http://archive.ics.uci.edu/ml/datasets/Poker+Hand>

Table 1. Collection Characteristics

Data Collection	Tree Height	#Inner nodes	#Leaf nodes	Index size [MB]	#Entries
Poker	5	3,176	35,345	78.9	1 000 000

In Table 2 we show performance of CPU and GPU algorithms. In the case of GPU we compare two columns. In first pass all accessed nodes are transferred to GPU. Although in second pass all nodes are already in GPU's memory. From results is obvious that time needed to transfer nodes to GPU's memory takes great portion of time. It takes at most $36\times$ more time than searching when all nodes are on GPU. The ratio between results of 1st and 2nd pass is decreasing as more queries is performed. This is happening because leaf nodes are transferred to GPU only first time they are accessed. From results we can see the CPU algorithm time is up to $4\times$ better than GPU. If we do not consider time needed to transfer nodes to GPU we see that GPU's algorithm is slightly faster only for low selectivity.

Table 2. R-tree Range Query Algorithms Results

Result Size	CPU [ms]	GPU (1st pass) [ms]	GPU (2nd pass) [ms]	AVG Accessed Leaf Nodes	AVG Accessed Inner Nodes
1	1.80	96.10	2.60	21.7	10.8
283.2	11.00	1338.50	37.10	338.7	1046.4
22 194.8	86.50	667.20	128.90	1214.7	8328.8
425 567.3	392.70	487.80	347.60	2582.0	27101.5

As we described in the previous chapter. GPU's algorithm cannot fill result set directly. In Table 3 we depicted performance of both algorithms without adding to the result set. The performance of single GPU algorithm is up $5\times$ better. Even in the case of higher selectivity it is $3\times$ more efficient.

Table 3. Comparison of Range Query Algorithms Without Adding to Result

Result Size	CPU Algorithm [ms]	GPU Algorithm [ms]	Speedup
1	0.00	0.01	0x
283.2	2.50	0.79	3x
22 194.8	26.50	6.27	4x
425 567.3	101.80	20.39	5x

4 Conclusion

In this article, we compared CPU and GPU variants of range query algorithms in the R-tree. We summarize the results. (1) The results of GPU shows the algorithm can be faster for low selectivity. However all accessed nodes must be stored in GPU's memory. (2) Performance of GPU algorithm is up to $5\times$ better if we do not consider data transfers between CPU and GPU and we do not fill result set. (3) Performance of GPU is negatively affected by relatively small amount of leaf nodes stored in buffer. In our previous work [1] we showed the GPU is more suitable for large data processing. (4) In the case of filling result set the GPU algorithm has a disadvantage because it cannot fill result set directly such as CPU.

As result, an integration into a DBMS kernel is not so straightforward and we must solve mainly these issues in our future work: the data transfer time must be effaced in the live cycle of DBMS and we must forward only low-selectivity range queries on a GPU because the number of compare operations increases for indices like R-tree in this case.

5 Projects and Publications

Pavel Bednář is a member of team solving the grant of GAČR, Czech Republic, No. SP2013/14. Pavel Bednář is a co-author of article [1].

Own References

1. P. Bednář, M. Krátký, P. Gajdoš, and P. Chovanec. Processing of Range Query Using SIMD and GPU. In *Proceedings of GID 2012 (GPUs in Databases), ADBIS 2012. To appear in AISC, Springer, 2012.*

Other References

2. R. Bayer and E. McCreight. Organization and Maintenance of Large Ordered Indexes. *Acta Informatica*, 3(1):173–189, 1972.
3. N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger. The R*-Tree: An Efficient and Robust Access Method for Points and Rectangles. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD 1990)*, 1990.
4. P. Chovanec and M. Krátký. Processing of Multidimensional Range Query Using SIMD Instructions. In *Informatics Engineering and Information Science*, volume 254. Springer, 2011.
5. M. Freeston. A General Solution of the n -dimensional B-tree Problem. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD 1995)*. ACM Press, 1995.
6. A. Guttman. R-Trees: A Dynamic Index Structure for Spatial Searching. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD 1984)*, pages 47–57. ACM Press, June 1984.

7. S. S. Lightstone, T. J. Teorey, and T. Nadeau. *Physical Database Design: the Database Professional's Guide*. Morgan Kaufmann, 2007.
8. nVIDIA. *Cuda Programming Guide*, 2012.
9. H. Samet. *Foundations of Multidimensional and Metric Data Structures*. Morgan Kaufmann, 2006.
10. M. Stonebraker, D. Abadi, A. Batkin, X. Chen, M. Cherniack, M. Ferreira, E. Lau, A. Lin, and S. Madden. C-store: A Column Oriented DBMS. In *Proceedings of the International Conference on Very Large Data Bases, VLDB'05*, 2005.

Polyaniline – constant electrical resistance

Jiří Bocheza and Stanislav Hejduk

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jiri.bocheza, stanislav.hejduk}@vsb.cz

Abstract. This article discusses the new findings in the area of research of the properties of conductive polymer - polyaniline (PANI). Specifically, it deals with new measurement results of changes of electrical resistance in dependence on the time when electrical current passes through.

1 Introduction

This article builds on already published research results of conductive polymer - polyaniline (PANI) in articles [1] and [2]. In article [1] a short description of properties of conducting polymers was carried out and the first set of basic measurements of individual potentially interesting optical, electrical or magnetic properties of this conducting polymer was evaluated. Article [2] dealt with the first results of measurement of change of electrical resistance of PANI layer in dependence on the time when electrical current passes through.

This article discusses other findings that were revealed by repeated and long-term measurements of electrical resistance of PANI layer in dependence on the time when electrical current passes through.

2 Connection diagram and method of measuring of electrical resistance of PANI layer in dependence on the time when electrical current passes through

Diagram in Figure 1 was used for measurement of changes of electrical resistance of PANI layer in dependence on the time when electrical current passes through.

Measuring devices were selected according to this diagram. For measurement the DC voltage power supply was used with voltage $U_{cc} = 10V$. The resistor $R_1 = 1,48M\Omega$ and resistor of PANI layer (R_{PANI}), which was applied to a laboratory micro-glass, were connected to this power supply. For connection of micro-glass with a PANI layer was used a holder, which was manufactured at the Department of Telecommunications at VŠB - TU Ostrava. Resistor R_1 was connected parallel with PC USB oscilloscope PicoScope 2204 from Pico Technology Company, which was connected to the PC. Measuring of a change voltage on the resistance R_1 in dependence on time was exe-

cuted with program PicoScope 6. The voltage and time from the measurements were automatically saved to a file for later processing.

The values of electrical resistance of PANI layer R_{PANI} were calculated using the derived formula: $R_{\text{PANI}} = (U_{\text{cc}} - U_1) / (U_1 / R_1)$.

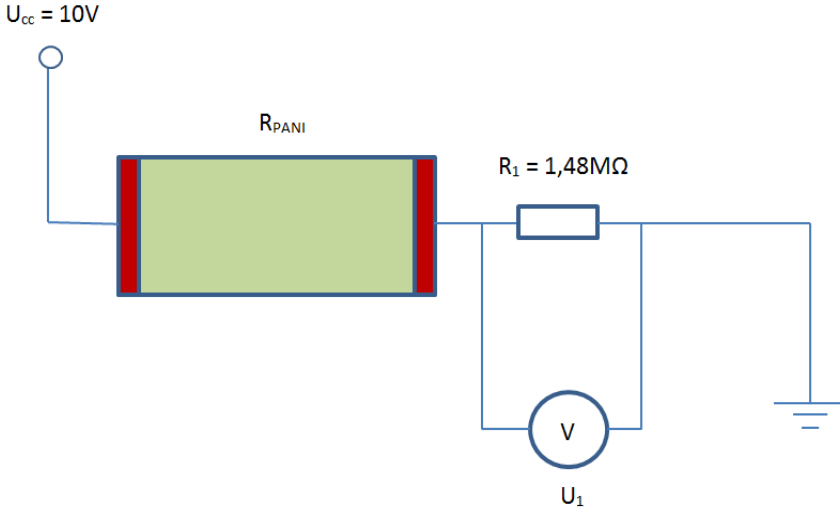


Fig. 1. Schema of measurement of changes of electrical resistance in depending on time when electrical current passes through.

3 Constant waveform of electrical resistance of PANI layer in dependence on the time when electrical current passes through

The measurements were performed on a sample of PANI, which was created in provisional conditions at the Department of Telecommunications at VŠB - TU Ostrava.

Before measurement, the sample was rinsed with faint hydrochloric acid and conductivity parameters of PANI layer were recovered [1], [3].

Then were measured waveforms of electrical resistance of PANI layer in dependence on the time when electrical current passes through. The time intervals were: 1 day, 1 month and 1 year after rinse in hydrochloric acid.

Measurements were performed by description in chapter 2. Electrical resistances of PANI layer were calculated from the measured values by the derived formula: $R_{\text{PANI}} = (U_{\text{cc}} - U_1) / (U_1 / R_1)$.

During these measurements we found that after some time it occurs in PANI layer to "steady" of electrical resistance in dependence on the time when electrical current passes through and this electrical resistance is constant. This measurement was repeated several times and it was verified that the waveform of electrical resistance was still constant.

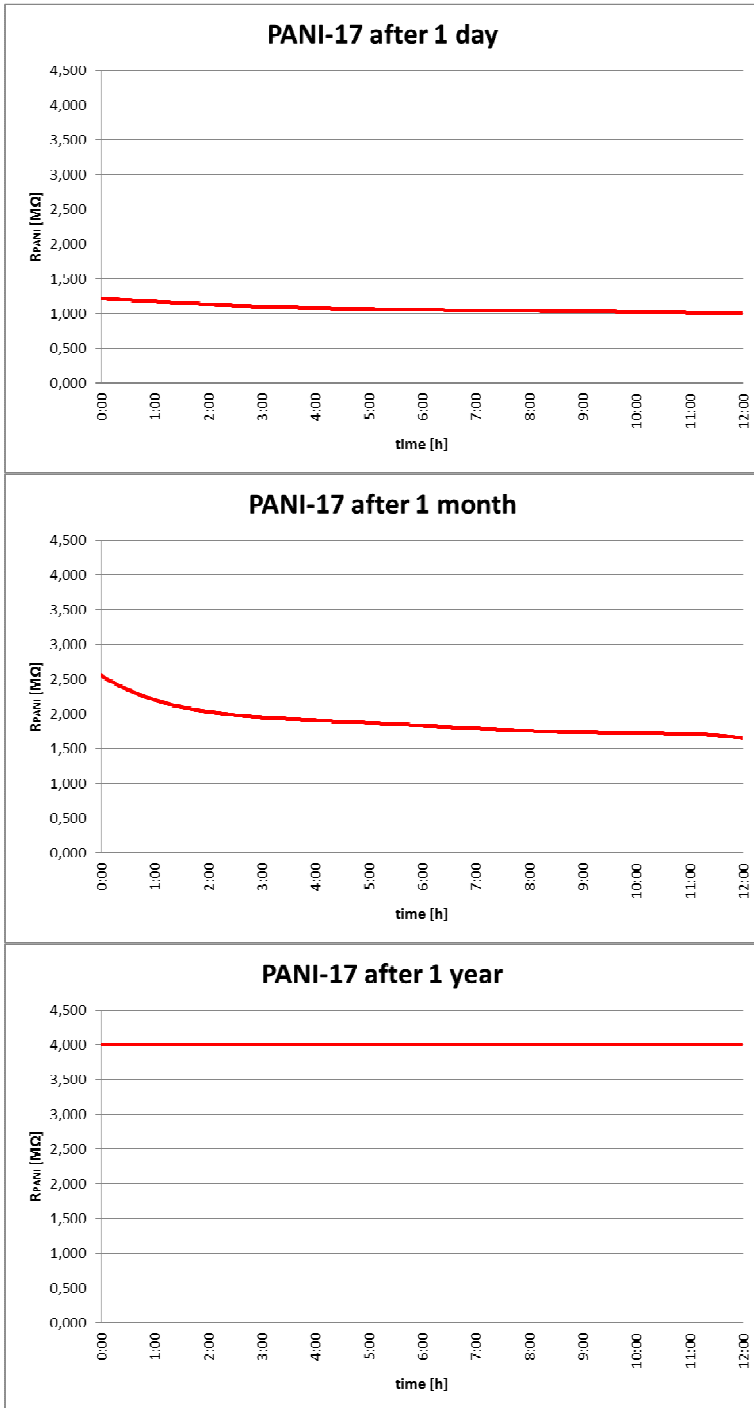


Fig. 2. Waveforms of changes of electrical resistance in dependence on time when electrical current passes through. Measurements were performed on one sample of PANI at time intervals: 1 day, 1 month and 1 year.

Figure 2 shows three graphs. The first graph shows the measured waveform of the change of electrical resistance of PANI layer in dependence on time when electrical current passes through. This waveform was measured one day after rinsing in hydrochloric acid and recovery of conductivity. The second graph shows waveform that was measured one month after rinsing and the third graph shows waveform that was measured one year after rinsing.

From waveforms, which were measured one day and one month after rinsing, it was previously thought that after some time after rinsing there is an increase of electrical resistance, but waveform stays roughly the same. From additional measurement is possible to say that this conclusion is half true. Still true that the electrical resistance of the PANI layer with increasing time after rinsing by hydrochloric acid is increased. This change is caused by the gradual conversion of conductive form (polyaniline salt) of PANI layer on the non-conductive form (polyaniline base) [1]. This transformation results in increase of electrical resistance. This transformation is spontaneous when the layer of PANI is exposed to the outside environment.

These additional measurements disprove that waveforms are similar and the electrical resistance increases only. From individual waveforms is evident that increasing time after rinsing leads to increase in electrical resistance. On the waveform, which was measured one year after rinsing, it is seen that over time the electrical resistance is stabilized and is constant when is exposed of electric current.

5 Conclusion

This article was focused on other measuring changes of electrical resistance of PANI layer during long-term effects of electric current.

From the measured waveforms (Figure 2) of the electrical resistance in dependence on the time when electrical current passes through we confirmed the assumption that with increasing time after rinsing of PANI layer in faint hydrochloric acid and restoring of conductivity parameters the electrical resistance of the PANI layer is increased, which is caused by the gradual conversion of conductive form (polyaniline salt) of PANI layer on the non-conductive form (polyaniline base). This transformation is spontaneous when the layer of PANI is exposed to the outside environment.

Another important finding is that over time after rinsing in hydrochloric acid leads to stabilization of electrical resistance when electrical current passes through and from the waveform, which was measured one year after rinsing, it is seen that the waveform of electric resistance is already constant in dependence on the time when electrical current passes through. We can say that the layer will be stable at a given time after rinsing and the resistance will be constant when electrical current will pass through. The question is, at what moment this condition occurs, and if it would be possible to speed up this time.

References

1. Bocheza, J., Hejduk, S., Vitásek, J. Conducting Polymers – Polyaniline. In *WOFEX 2011 : 9th annual workshop: proceedings of papers*. 1st edition. Ostrava : VSB - Technical University of Ostrava, 2011.
2. Bocheza, J., Hejduk, S. Polyaniline – electrical resistance. In *WOFEX 2012 : 10th annual workshop: proceedings of papers*. 1st edition. Ostrava : VSB - Technical University of Ostrava, 2012.
3. Stejskal, J., Sapurina, I., Prokeš, J., Zemek, J.: In-situ polymerized polyaniline films. *Synthetic Metals*. 1999, 105, pp. 195-202. ISSN 0379-6779.

Energy Based Descriptors and their Application for Object Detection

Radovan Fusek, Eduard Sojka, Karel Mozdřeň, and Milan Šurkala

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{radovan.fusek, eduard.sojka, karel.mozdren, milan.surkala}@vsb.cz

Abstract. In this paper, we propose a novel technique for object description. The main idea of the proposed technique is based on the investigation of energy distribution (in the image) that describes the appearance of objects. The distribution is encoded into a vector of feature. The vector is then used as an input for the SVM classifier. The technique can be used for detecting arbitrary objects. In this paper, we use it for detecting the cars, pedestrians, and faces. Compared with the state-of-the-art descriptors that are based on the histogram of oriented gradients, the proposed approach achieved better results, especially from the viewpoint of dimensionality of the feature vector; our approach is able to successfully describe the properties of objects with a relatively small set of numbers.

1 Introduction

In feature-based detectors, the selection of relevant features that are able to reliably describe the objects of interest is a key point. In the recent years, the object detectors that are based on the edge analysis that provide the valuable information about the objects of interest have been used in many detection tasks. In this area, the Histograms of Oriented Gradients (HOG) [3] are considered as the state-of-the-art method. In HOG, a sliding window is used for recognition. In the process of obtaining HOG descriptors, the window is divided into small connected cells. The histograms of gradients are calculated for each cell. It is desirable to normalize the histograms across a large block of image. As a result, a vector of values is computed for each position of window. This vector is then used for recognition, e.g. by the Support Vector Machine (SVM) classifier. The HOG descriptors are very useful in many detection tasks. F. Suard [6] et al. proposed the method for pedestrian detection using the HOG descriptors with the SVM classifier. Boosting HOG features for the vehicle detection in airborne videos are presented in [1].

In contrast with the HOG descriptors, the proposed method captures the object information by the energy distribution of object areas instead of the distribution of gradient magnitudes and directions. Using the proposed method, the feature vectors of relatively small dimensions are sufficient for successful description of the objects. The process of extraction of the proposed features is described in Section 2, the results are shown in Section 3, 4, and 5.

2 Proposed Method

In the process of extracting the proposed features, the image inside the sliding window is divided into the regular blocks. We use the gravity centers of these blocks as the places in which we put the temperature sources. For the purpose of obtaining the distribution of temperature, the blocks are divided into the small connected cells.

Let $I(x, y, t)$ be a value of temperature at a time t and at a position (x, y) . Inside each cell, the mean temperature $I_{\mu_{it}}$ of the i -th cell at a time t can be calculated. The final feature vector is composed of these mean values (Fig. 1). We note that the temperature transfer is computed in the whole image inside the sliding window and temperature transferred from one source can influence every cell inside the sliding window; the blocks and cells are formed only for distribution measurement. In the next step, the SVM classifier is trained over the proposed descriptors.

For the practical realization of the method, it is important to mention that the thermal field over one position of sliding window can be solved by making use of the following equation [5].

$$\frac{\partial I(x, y, t)}{\partial t} = \text{div}(c\nabla I), \quad (1)$$

where I represents the temperature at a position (x, y) and at a time t , div is a divergence operator, ∇I is the temperature gradient and c stands for thermal conductivity. For the source points and arbitrary time $t \in [0, \infty)$, we set $I(x_s, y_s, t) = 1$, where (x_s, y_s) are the coordinates of source points (i.e. we hold the temperature constant during the whole process of transfer, which is in contrast with the usual diffusion approaches). In all remaining points, we take into account the initial condition $I(x, y, 0) = 0$. We solve the equation iteratively.

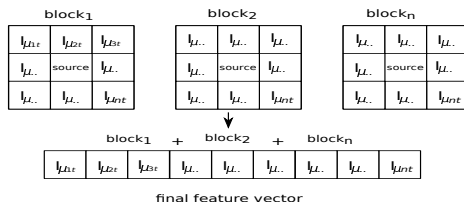


Fig. 1. The feature vector that is composed of the mean temperatures of cells.

3 Pedestrian detection

For this task, we collected 2500 positive samples and 10000 negative samples for the training phase. For the positive set, we combined the pedestrian images from

the CBCL Pedestrian Database [2] with the images from the Daimler benchmark [4]. For the negative images, the examples were randomly sampled from the INRIA Person Dataset [3]. For the proposed method, each sample was resized to the size of 91×151 pixels.

We experimented with the parameters of proposed features and we suggested the $Energy_{480}$ configurations. The $Energy_{480}$ configuration was designed with the size of block = 15×15 pixels; size of cells = 5×5 pixels (the number of cells inside each block = 8); size of sources = 5×5 pixels; time for the temperature transfer = 100 (the number of iterations). This configuration consists of 480 descriptors for one position of sliding window. For the comparison, we designed the HOG_{3780} configurations of classical HOG features. The HOG_{3780} configuration was designed with the typical parameters of HOG descriptors; the size of block = 16×16 pixels; size of cell = 8×8 pixels; horizontal step size = 8 pixels; number of bins = 9. This configuration consists of 3780 HOG descriptors for one position of sliding window. For the testing, we collected 55 images from the [3]. The accuracy results of approaches are shown in Table 1. The detection results of the $Energy_{480}$ configuration are shown in Fig. 2.

Table 1. The detection performance (pedestrian detection).

	Precision	Sensitivity	F1 score
$Energy_{480}$	87.14%	84.72%	86.56%
HOG_{3780}	86.97%	81.97%	84.03%

Compared with the HOG_{3780} configuration (F1 score 84.03%), the proposed features achieved the similar results, however the proposed method gives $7 \times$ less descriptors than the classical HOG_{3780} configuration.

4 Car detection

For the training phase, we collected the data set consisting of 900 samples (450 vacant spaces/non-cars, 450 occupied spaces/cars). Each sample was resized to the size of 90×90 pixels. We designed the following configuration of our descriptors that is denoted as $Energy_{144}$ with the size of block = 15×15 pixels, size of temperature sources = 3×3 pixels, number of cells inside the block = 4, iterations (time) for the temperature transfer = 450. This configuration consists of 144 descriptors for one position of sliding window. For comparison, we created three types of HOG descriptors HOG_{1296} , HOG_{900} and HOG_{300} . These configurations consist of 1296, 900, and 300 HOG descriptors for one position of sliding window.

To calculate the performance of approaches, we used Matthew’s correlation coefficient (MCC) that is typically used in machine learning to assess the performance of a binary classifier and MCC is useful when two classes have a different size. (In our case, the numbers of TP, TN, FP, FN are different). The values

of MCC are between -1 and +1. The higher value represents better predictions. We collected 28 testing images from the parking lot to evaluate approaches. The testing images were not used in the training phase and the images were taken in several year seasons (in different weather and lighting conditions).

Table 2. The detection performance (occupancy detection).

	sunny					winter/rain/night					overall performance
	TP	TN	FP	FN	MCC	TP	TN	FP	FN	MCC	MCC
HOG_{300}	406	351	25	2	0.93	148	457	170	9	0.55	0.76
HOG_{900}	408	353	23	0	0.94	156	512	115	1	0.68	0.83
HOG_{1296}	408	357	19	0	0.95	156	317	310	1	0.41	0.65
$Energy_{144}$	408	361	15	0	0.96	146	599	30	9	0.85	0.93

The proposed method shows that the cars can be described with a reasonable number of features with very good detection results without need for the methods for reducing the feature space. The accuracy results of approaches are shown in Table 2. The example of detection results of our approach is shown in Fig. 2.

5 Face detection

For the training phase, our positive set consists of 1700 faces. We used the face images from the BIOD database combined with the Caltech face dataset. We manually cropped these images on the area of faces only. The negative set consists of 3000 images that was obtained from the MIT-CBCL database. We resized all training images to the size of 90×90 pixels. We experimented with the parameters of our method and we suggested the following configuration: ETF_{324} . The ETF_{324} configuration is designed with the size of block 5×5 , time chosen for the transfer of temperature $t = 200$, size of temperature sources is 1×1 pixels, number of cells inside the block is 1.

For comparison, we used the detectors that are based on the HOG features, LBP (Local Binary Patterns) features and Haar features (Viola-Jones detection framework); the HOG_{324} configuration was designed with the same number of feature values like in the ETF_{324} configuration. The parameters were as follows: the size of block = 32×32 , size of cell = 16×16 , horizontal step size = 32, number of bins = 9. This configuration gives 324 HOG features. To calculate the performance of approaches, we collected the set of 80 images that contains 117 faces from the Faces in the Wild dataset. In Table 3 and Fig. 2, the detection results are shown.

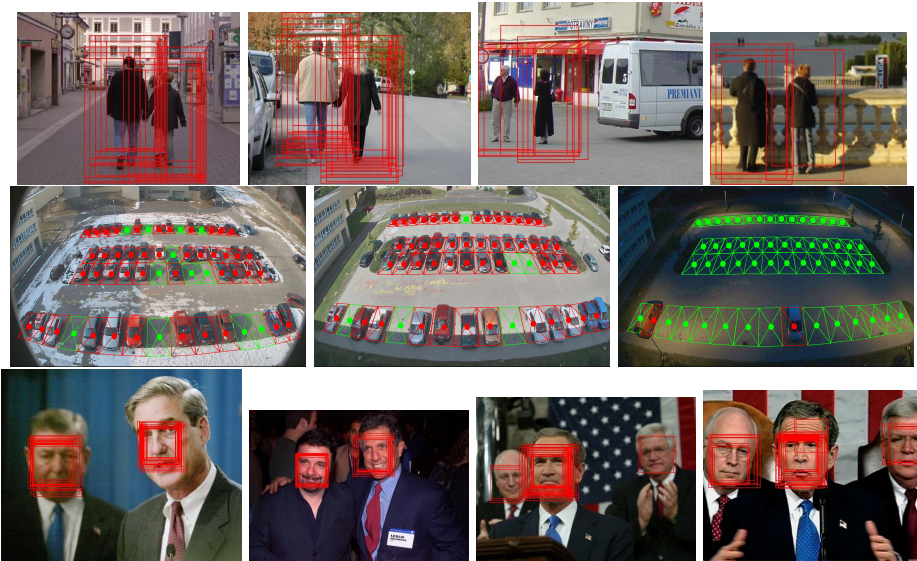
The HOG based detector (HOG_{324}) is not able to describe all details of the faces with such a small amount of descriptors (F1 score = 79.89%). We experimented with many configurations of HOG features, but increasing the number of features (without increasing training data and without the reduction of feature space) did not improve the detection accuracy. Haar based detector

Table 3. The detection performance (face detection).

	Precision	Sensitivity	F1 score
ETF_{324}	87.90%	92.32%	90.08%
HOG_{324}	67.25%	98.29%	79.89%
$Haar$	74.45%	87.18%	80.31%
LBP	62.07%	76.92%	68.70%

had less false positive detections (precision = 74.45%) than the configuration of HOG features and LBP based detector. However, the Haar based detector and LBP based detector missed some of the faces (sensitivity = 87.18% and 76.92%, respectively).

Finally, the proposed technique shows that the faces, pedestrians and cars can be described with a reasonable number of features with very good detection results and also with a relatively small set of training data without need for the methods for reducing the feature space.

**Fig. 2.** The detection results of our approach.

6 Conclusion

In this paper, we proposed the novel approach for the computation of image features that are based on the distribution of energy (temperature). We demon-

strated that the features were able to describe the faces, cars, and pedestrians with a relatively small number of relevant features and with the high accuracy.

Acknowledgments

This work was supported by the SGS in VSB Technical University of Ostrava, Czech Republic, under the grant No. SP2013/185.

Author's Publications

The parts of results were presented in:

- Fusek, R., Mozdren, K., Surkala, M., Sojka, E.: **AdaBoost for Parking Lot Occupation Detection**. In Proceedings of the 8th International Conference on Computer Recognition Systems, CORES 2013, pp. 681-690 (2013)
- Fusek, R., Sojka, E., Mozdren, K., Surkala, M.: **Energy-Transfer Features and their Application in the Task of Face Detection**. In Proceedings of the 10th International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2013, to appear (2013)
- Fusek, R., Sojka, E., Mozdren, K., Surkala, M.: **Energy-Transfer Features for Pedestrian Detection**. In Proceedings of the 8th International Symposium on Visual Computing, ISVC 2013, to appear (2013)

References

1. Xianbin Cao, Changxia Wu, Pingkun Yan, and Xuelong Li. Linear svm classification using boosting hog features for vehicle detection in low-altitude airborne videos. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 2421–2424, sept. 2011.
2. Center for Biological and Computational Learning. MIT CBCL Pedestrian Database #1, 2013. <http://cbcl.mit.edu/software-datasets/PedestrianData.html>.
3. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893 vol. 1, june 2005.
4. M. Enzweiler and D.M. Gavrilu. Monocular pedestrian detection: Survey and experiments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2179–2195, 2009.
5. P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12:629–639, July 1990.
6. F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi. Pedestrian detection using infrared images and histograms of oriented gradients. In *Intelligent Vehicles Symposium, 2006 IEEE*, pages 206–212, 0-0 2006.
7. Qiang Zhu, Mei-Chen Yeh, Kwang-Ting Cheng, and S. Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1491–1498, 2006.

Image Segmentation via Graph Cuts using Residual Graph

Michael Holuša

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
michael.holusa@vsb.cz

Abstract. In this paper, we present a new image segmentation method using iterated graph cuts. In the standard graph cuts method, the data term is computed on the basis of the color distribution of object and background. In this case, some background regions with the color similar to the object may be incorrectly labeled as an object. We try to overcome this drawback by introducing a new data term that reduces the importance of color distribution. This reduction is realised by a new part that uses data from a residual graph that remains after performing the max-flow algorithm. According to the residual weights, we change the weights of t-links in the graph and find a new cut on this graph. The results and comparison with other graph cuts methods are presented.

1 Introduction

The interactive graph cuts image segmentation method was proposed in 2001 by Boykov and Jolly [1]. They presented a new method for binary segmentation, where the image is divided into two segments - foreground and background. GrabCut segmentation method [2] replaces the single graph cuts by an iterative procedure. The data term in both methods is constructed on the basis of a priori known color distribution of object and background. This approach need not to work in images in which the color of object and background is similar. In our method, we decrease the importance of the color distribution and add a new part that is independent on color, but uses data from a residual graph. We change the weights in the original graph and find new cuts repeatedly.

The paper is organised as follows. A review of standard graph cuts method is presented in Section 2. Section 3 contains a description of our iterative method. The experiments and a comparison with other graph cuts methods are in Section 4. Section 5 is a conclusion.

2 Graph Cuts in Image Segmentation

In this section, we briefly describe an image segmentation method based on the graph cuts. Let \mathcal{P} be a set of all image points, let \mathcal{N} be a set of all neighboring points, and let \mathcal{L} be a set of possible labels. The goal is to find a labeling

$l = \{l_p | l_p \in \mathcal{L}\}$ that assigns a label $l_p \in \mathcal{L}$ to each image pixel $p \in \mathcal{P}$ such that this labeling minimizes the energy function

$$E(l) = \sum_{p \in \mathcal{P}} R_p(l_p) + \lambda \sum_{(p,q) \in \mathcal{N}} B_{p,q} l_p, l_q. \quad (1)$$

The first term of this function is called the data term. It sets individual penalties for assigning a label l_p to a pixel p . The second term, called the smoothness term, penalizes the discontinuity of labeling between the neighboring pixels, l_p and l_q stand for a labeling of points p and q , respectively. The coefficient λ sets a relative importance of the smoothness term versus the data term. The data term is defined as negative log-likelihoods

$$R_p(\text{"obj"}) = -\ln \Pr(I_p | \mathcal{O}), \quad (2)$$

$$R_p(\text{"bkg"}) = -\ln \Pr(I_p | \mathcal{B}), \quad (3)$$

where \mathcal{O} , \mathcal{B} represent the object and background points, respectively, I_p is an intensity of a given pixel. The smoothness term is defined as

$$B_{p,q} \propto \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right). \quad (4)$$

The graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a set of vertices \mathcal{V} and a set of edges \mathcal{E} . The set $\mathcal{V} = \mathcal{P} \cup \{S, T\}$ contains all nodes from \mathcal{P} and two additional terminal nodes S and T representing the source (foreground) and the sink (background), respectively. The set \mathcal{E} contains two types of edges: t-links connect the nodes from \mathcal{P} with both terminals, and n-links connect the neighboring non-terminal nodes. The weights of t-links correspond to the Eqs. (2) and (3). The weights of n-links are set on the basis of Eq.(4). For the minimization, we used the algorithm from [3]. More information about the method can be found in [3–5].

3 Iterated Graph Cuts with Residual Graph

In the original graph cuts, the likelihoods in Eqs. (2) and (3) are computed only from the seeds. If the seeds are marked ineptly, probably the final segmentation, carried out by the one-shot graph cut, will be incorrect. This drawback had been overcome by the iterative approach [2], where the solution is reached by a sequence of graph cuts. When the min cut is found, this estimated segmentation is used to specify the color distribution of foreground and background, the weights of t-links in the original graph are changed, and the next cuts are carried out. Using the color distribution seems as a logical approach that works in many situations but, in some images, background areas have a color similar to the object. Then the methods may label some non-object points as the object incorrectly. In our solution, we propose a new part of the data term that is not based only on the color distribution. The construction of this data term is described in the following subsections.

3.1 Our Data Term

Our new data term consists of two parts. Use the notations \mathcal{O} and \mathcal{B} for the sets of object and background points, respectively. The first part is similar to Eqs. (2) and (3). The likelihoods $\Pr(I_p|\mathcal{O})$ and $\Pr(I_p|\mathcal{B})$ are transformed to posterior probabilities $\Pr(\mathcal{O}|I_p)$ and $\Pr(\mathcal{B}|I_p)$. It causes that the probabilities are summed up to 1, which is a more suitable form for our data term, because all the likelihoods will have similar importance. In the second part of the data term, we assume that we have no information about the color distribution of object and background. Since we want to minimize the false object detections, we decrease the probability of being object point for every image point except the seed points ($\Pr(\mathcal{O}) < \Pr(\mathcal{B})$). Let ξ be the coefficient that decreases the probability. Then

$$\Pr(\mathcal{O}) = 0.5 - \xi, \quad (5)$$

$$\Pr(\mathcal{B}) = 0.5 + \xi. \quad (6)$$

For the object seeds, we have $\Pr(\mathcal{O}) = 1$, $\Pr(\mathcal{B}) = 0$, and vice versa for the background seeds. At this point, all non-seeded points suffer from a reduction of probability, although many of them, in fact, should belong to the object. Our goal is to return the reduced value back for these points that are mainly located in the neighborhood of the detected object areas. We found out that the residual graph, described in the next subsection, has promising properties for determining these points. The final segmentation is found iteratively because the segmentation result from the previous process is used for setting the weights of t-links in the next step.

3.2 Residual Graph

Let \mathcal{G}_{res} be a residual graph, i.e., a graph that remains after executing the max-flow algorithm. Since the data term in Eq. (1) is related to the t-links, we will focus only on the t-links of residual graph. If the max-flow is reached, at least one t-link is saturated for each $p \in \mathcal{P}$ (otherwise, there is a path between the terminals). The second t-link has some residual capacity or is also saturated. In Eq. (5), we decreased the probability of being object for all image points, except the seeds. Due to this fact, we need to decide whether some of these points should not be possibly labeled as object (and return the reduced probability back). Let $w_{\text{res}}(p, T)$ be the residual capacity of the t-link connecting p and the sink terminal T . This capacity says, how “far” the edge is from saturating. If the edge is saturated or its weight is very low, there is a certain chance that the point should belong to the object instead the background as it was originally detected. If, on the other hand, the value of $w_{\text{res}}(p, T)$ is high, p was marked strongly (and probably correctly) as a background point. We define a new term that reflects the properties of residual weights. For the t-link with a low residual weight, we correct the probability more than for the t-link with a high residual

weight. Let k stands for the iteration step. The residual term f_{res} is of the form

$$f_{\text{res}}(p)^{(k)} = \exp\left(\frac{-w_{\text{res}}^{(k-1)}(p, T)}{2\sigma^2}\right), \quad (7)$$

where σ estimates how low residual weight of edge $\{p, T\}$ allows to the point p to be alternatively considered as object point in the next step. The t-links with low residual weights are mainly located around the areas convincingly detected as object points. It is caused by a flow from the object terminal; the flow is transported through the t-links $\{S, p\}$ mainly to the neighboring object points. Since the t-link $\{S, p\}$ of the object point has a high weight and the second t-link $\{p, T\}$ has a very low weight, the only possible flow is through the points neighboring with the object points. Then the flow reaches the sink through the t-links of neighboring points. It causes that, for the neighboring points, the weights of t-links between the point and the sink terminal are decreasing; in many cases, the edges are saturated. The flow to the points outside the object is eliminated by a low weight of the n-links connecting the points with a higher color difference, which usually signalizes an edge between the object and background. The behaviour of residual graph can be likened to a spreading of object points from the seeds to the neighboring areas.

3.3 Iterated Graph Cuts

According to the properties of residual graph, we modify the weights of the t-links in the original graph and find the new cut. The proposed method certainly labels only new object points, not the background points. If a point is labeled as an object in any iteration, it is added to the set \mathcal{O} . All other points, except the background seeds, are labeled as a background, but may be labeled as an object in next iterations. To label certain background points too, we add one more parameter into our data term. We assume that if a point is not labeled in the k -th iteration as an object, the probability that it will change in the $(k+1)$ -th iteration is lower. Then, for all the points $p \in \mathcal{P}, p \notin (\mathcal{O} \cup \mathcal{B})^{(k)}$ the inequality $\Pr(\mathcal{O})^{(k)} > \Pr(\mathcal{O})^{(k+1)}$ holds. We propose the coefficient that decreases the probability in the form of $(ck)^2$, where c sets how fast the value grows with respect to k . The square causes that it grows slower in the first steps when a change of object is more probable. The value of c depends on the complexity of object. If the object has constant color, c can be higher, and the method converges faster. All the previously described features lead to the definition of our new data term as

$$f_{\mathcal{O}}(p)^{(k)} = \eta \Pr(\mathcal{O}|I_p) + (1 - \eta) (\Pr(\mathcal{O}) + \xi f_{\text{res}}(p)^{(k)} - (ck)^2), \quad (8)$$

which can be understood as a linear combination of our new part and the posterior probability based on the color of pixel; as an initial value, we take $w_{\text{res}}^0(p, T) = \infty, \forall p \in \mathcal{P}$. The new part consists of the reduced prior probability (Eq. (5)), the residual term (Eq. (7)) and the decreasing coefficient that reduces

the probability during the iterations. If the probability for any point p is lower than a chosen threshold (close to zero), p is considered to be a background point and is added to the set \mathcal{B} . The η parameter sets the importance of the term relating to the color distribution. Since the probabilities $\Pr(\mathcal{O}|I_p)$, $\Pr(\mathcal{B}|I_p)$ and $\Pr(\mathcal{O})$, $\Pr(\mathcal{B})$ are summed up to 1, and the value of $f_{\mathcal{O}}(p)$ does not leave the interval $\langle 0, 1 \rangle$, the penalties from Eq. (2) and Eq. (3) are replaced by

$$R_p(\text{“obj”})^{(k)} = -\ln f_{\mathcal{O}}(p)^{(k)}, \quad (9)$$

$$R_p(\text{“bkg”})^{(k)} = -\ln (1 - f_{\mathcal{O}}(p)^{(k)}). \quad (10)$$

In the first step, the sets \mathcal{O} and \mathcal{B} contains only seeds of object and background, respectively. If the set \mathcal{O} or \mathcal{B} changes in any iteration, the likelihoods $\Pr(\mathcal{O}|I_p)$ and $\Pr(\mathcal{B}|I_p)$ are updated. After a certain number of steps, all the points will be in the set \mathcal{O} or \mathcal{B} . A point p is added to \mathcal{O} if it is labeled as object in any of iterations, and it is added to \mathcal{B} , if the value of $f_{\mathcal{O}}(p)$ is close to zero. The method converges to the state in which every point p will have one t-link with zero capacity.

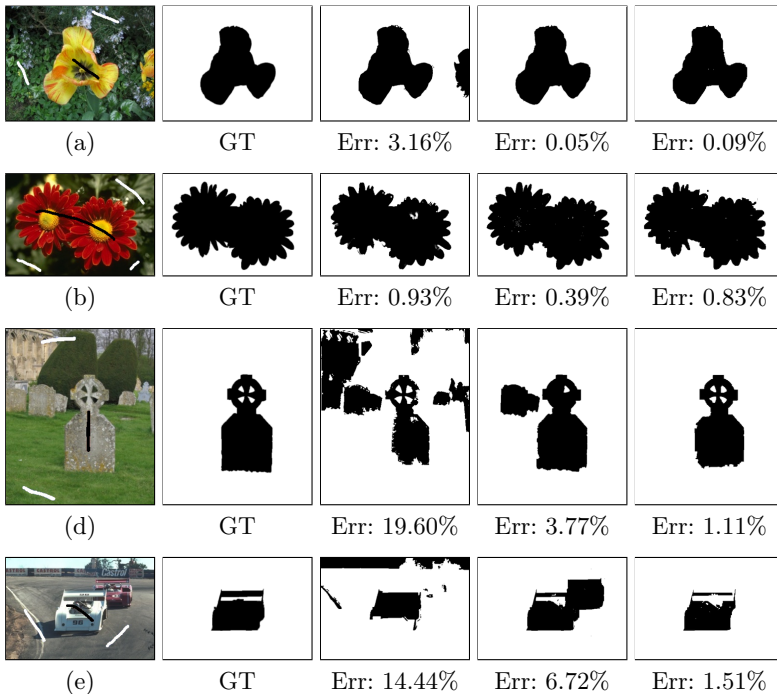


Fig. 1. The results of segmentation provided by the standard graph cuts [1] (column 3), GrabCut [2] (column 4) and our method (column 5). The strokes in the input images (column 1) represent the seeds. The ground truth for each segmentation is in the column 2. Below each result, there is the error rate of each segmentation.

4 Experiments

In this section, the results of our method are validated and compared with the standard graph cuts method [1] and with GrabCut [2]. The results of the new method and the comparisons are in Fig. 1. The error rate is computed as the ratio of number of points that differ between segmentation and the ground truth to the number of pixels in the whole image. Identical seeds are used for all the compared methods. In GrabCut, the color model of background is constructed via rectangle around the desired object. In our experiments, we put the rectangle along the border of image and, in addition, the seeds are used as well. According to the error rates in Fig. 1, our method outperforms the standard graph cuts algorithm. Comparing to GrabCut, our method is slightly worse, if the object is unique (Fig. 1(a), 1(b)). On the other hand, if the image contains more objects or background areas with the color similar to the object, our method provides better results (Fig. 1(c), 1(d)).

5 Conclusion

We have presented a new segmentation method using the graph cuts algorithm. We reduced the importance of color distribution, which is typical for previously presented graph cuts methods, and added a new part into the data term that consists of prior probability in combination with data from a residual graph. Our approach eliminates false object detection that frequently appears in other methods if the colors of foreground and background are similar. The graph cuts are carried out iteratively until the energy converges. Our method achieves better results than the standard graph cuts and it is slightly worse than GrabCut if the image contains a unique object. On the other hand, our method achieves better results if the image contains background with a color similar to the object color.

This work is a part of the paper published in the conference proceedings at the 9th International Symposium on Visual Computing.

References

1. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In: Computer Vision, 2001. ICCV 2001. Proceedings. 8th IEEE International Conference. (2001) 105–112 vol.1
2. Rother, C., Kolmogorov, V., Blake, A.: "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **23** (2004) 309–314
3. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **26** (2004) 1124–1137
4. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient n-d image segmentation. *Int. J. Comput. Vision* **70** (2006) 109–131
5. Boykov, Y., Veksler, O.: Graph Cuts in Vision and Graphics: Theories and Applications. In: Handbook of Mathematical Models in Computer Vision. Springer US (2006) 79–96

Indoor localization system using WiFi fingerprinting and built-in inertial sensors

Jakub Jirka

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jakub.jirka1@vsb.cz

Abstract. This paper is concerned with application for position localization using smartphone built-in wireless adapter and inertial sensors. Wireless adapter is used to determine initial device position and inertial sensors like accelerometer, magnetic sensor and gyroscope are with Newton's laws used to determine device movement. Conjunction of these two approaches brings better results than the standard triangulation or trilateration algorithms.

1 Introduction

Smartphone devices are spreading more and more in this society [1] and these devices themselves contain one of the most sophisticated sensors able to detect user motion in the three dimensional space we are living in. The application itself uses two types of “sensors”:

1. Wireless adapter for initial position estimation
2. Inertial sensors used for motion estimation

This paper describes use of these two sensor groups and shows measured results and tests.

Wireless adapter is using fingerprinting algorithm for initial position detection in known space.

Built-in inertial sensors which are:

- ST Microelectronics accelerometer - LIS331DLH capable of measuring acceleration $\pm 8g$ on 16bit ADC. Thus accelerometer gives sensitivity of 3,9mg/digit. Sampling frequency of this accelerometer varies from 23-26Hz in the normal setting but can go up to 50Hz in case of necessity.
- Asahi Kasei Microdevices AK8973 3-axis Magnetic field sensor with sensitivity 0,0625 μ T/digit. This sensor contains Hall's sensors and thermometer for error correction.

Following chapters describe algorithms used for position estimation and tests.

2 Measurement

2.1 Wireless adapter fingerprinting

Initial position is estimated using fingerprinting algorithm [3]. Every smartphone device has wireless adapter that scans surrounding area for wireless AP (Access Point) routers and returns following basic set of information about every wireless AP:

- SSID (Service Set Identifier) – human readable name of the wireless network
- BSSID (Basic Service Set Identifier) – wireless network MAC address
- RSSI (Received Signal Strength Intensity) – received signal intensity on the device’s antenna in dBm units expressing power ratio relative to 1mW of power.

With fingerprinting algorithm known area is “fingerprinted” which means that set of measurements takes place in the know area in four 90° directions. Such a map is shown on picture figure (Fig. 1).

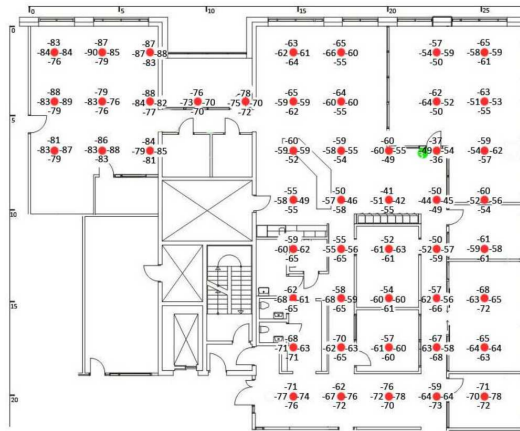


Fig. 1 Fingerprinted area

All these measurements (RSSI values, x and y positions) are stored in the database. Values in this database are compared to the localized device measurement somewhere in the known space. The closest position estimation is calculated using Euclidian distance algorithm from RSSI values which are refreshed every 500ms.

Euclidian distance

Euclidian distance expresses distance between two points in space using Pythagoras theorem. Euclidian distance between two points is then given according to equation (1).

$$d(p,q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} = \sum_{i=1}^n (q_i - p_i)^2 \quad (1)$$

- q**...measured point
- p**...reference point
- n**...number of access points

X and Y coordinates are then calculated using weighted sum algorithm of coordinate's mean value from the **m** number of the closest access points, equations (2, 3).

$$X = \sum_{i=1}^m \left(\frac{1}{d_i}\right)^2 * \left[\frac{x_1}{d_1} + \frac{x_2}{d_2} + \dots + \frac{x_m}{d_m} \right] \quad (2)$$

$$Y = \sum_{i=1}^m \left(\frac{1}{d_i}\right)^2 * \left[\frac{y_1}{d_1} + \frac{y_2}{d_2} + \dots + \frac{y_m}{d_m} \right] \quad (3)$$

- x_i** and **y_i**...coordinates of selected fingerprints
- d**...distance to every node counted by Euclidian distance algorithm
- m**...number of selected nodes

Selected **m** numbers of nodes are the ones with the smallest value of **d** (closest ones to the currently measured position).

According to the fact that RSSI signal rapidly fluctuates in time, as seen on the picture (Fig. 2) set of correction data from built-in inertial sensors is used.

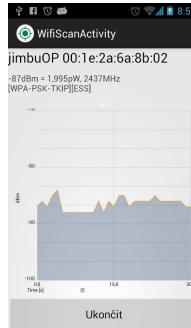


Fig. 2 Measured wireless signal

2.2 Inertial sensors

Inaccurate and time fluctuating wireless position data are corrected during located device movement using built-in inertial sensors like accelerometer, magnetic sensor and gyroscope.

Speed of the movement is calculated from linear acceleration and direction is taken from magnetic sensor that serves as a compass. To avoid linear accelerations along the directions that does not introduce forward or reverse movement an algorithm detecting the user movement (step detection algorithm) and filter out the false shaking or other stationary accelerations.

Linear acceleration

Linear acceleration which is used to calculate walked distance is obtained from accelerometer and magnetic sensor. These sensors values are used to get the device's orientation and gravitational pull along X, Y and Z axes.

The most important acceleration value is longitudinal acceleration which is expressed as the acceleration along the Y axis (In case device is parallel with the earth's plane).

The movement's velocity (Fig. 3) can be calculated from this acceleration with equation (4).

$$d(p,q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} = \sum_{i=1}^n (q_i - p_i)^2 \quad (4)$$

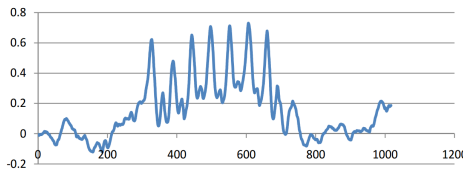


Fig. 3. Velocity time function calculated from acceleration

The movement's walked distance (Fig. 4) can be calculated according to equation (5).

$$d_n = \sum_{i=1}^n (v_i * \Delta t + \frac{1}{2} a_i \Delta t^2) \quad (5)$$

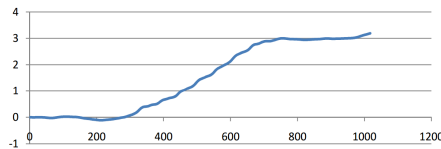


Fig. 4. Walked distance calculated from velocity

Measured values were approximately 3x smaller than real walked distance. This error was caused by the false acceleration along the X axis when the phone is rotating and moving along its own axis. This error is fixed with the gravitational pull correction calculations.

The problematic describing's phone orientation and gravitation components influence is vast and is not described in this article. Equation used for the corrected calculations is equation (6), where G_{total} is total component summation from all gravitational pull components.

$$Lin_{Acc_y} = Acc_y - G_{total} * \cos\left(\frac{\pi}{2} + \theta_y\right) \quad (6)$$

2.3 Magnetic sensor and orientation

To define device's direction while walking device's magnetic sensor is used with the accelerometer. Accelerometer is used to determine device's own initial orientation. For this orientation calculation gyroscope with its angular movement data can be used as well. Angular movement can be integrated over time summed and the result would be the angle change of the device.

This application uses magnetic sensor to determine device's azimuth (Fig. 5) (0° - north, 90° - east, 180° - south, 270° - west).

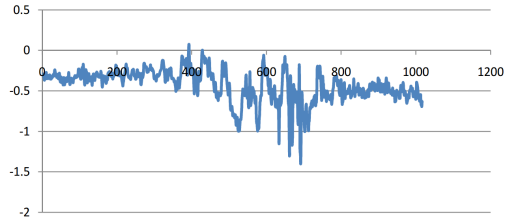


Fig. 5. Device's azimuth time function.

2.4 Test and results evaluation

The tests were divided and analyzed the two major parts of the algorithm the wireless fingerprinting algorithm and inertial built-in sensors movement detection algorithm.

Wireless adapter fingerprinting algorithm tests

Algorithm is using time moving averaged data calculated from the wireless adapter RSSI values. The maximum error during this testing was approximately 4m (Fig. 6).

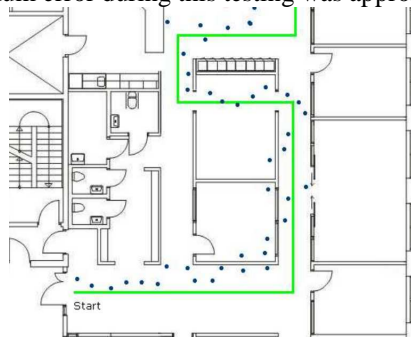


Fig. 6. Wireless fingerprinting algorithm test

Built-in inertial sensors movement detection tests

The device was tested in the building where the test had 15 steps and 7,8 meters were walked during these 15 steps. Calculated result of walked distance is seen on figure (Fig. 7). The overall result of walking can be seen on (Fig. 8).

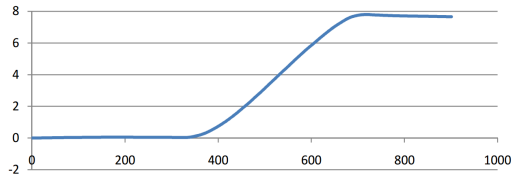


Fig. 7. Test's walked distance time function

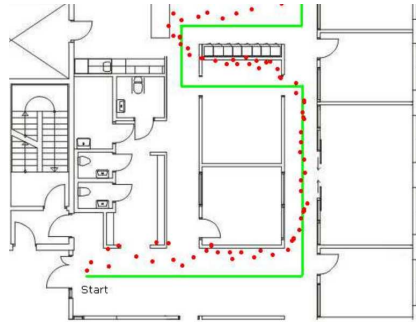


Fig. 8. Walking test with the built-in inertial sensors

Application implementation notes

Application was implemented on android device using purely java programming language. Algorithms testing and evaluation was performed in the python programming language.

Conclusion

This article proposes improvement for current fingerprinting based localization algorithms using built-in inertial sensors to improve movement detection algorithms.

Tests taken on this algorithm proved improvement from the solely wireless adapter based localization algorithms.

References

1. M. Brownlow. "Smartphone statistics and market share", Email marketing reports, June 2010
2. Jakub Jirka and Michal Prauzek. Built-in smartphone accelerometer motion pattern recognition using wavelet transform. International Joint Conference CISIS'12-ICEUTE'12-SOCO'12 Special Sessions Advances in Intelligent Systems and Computing Volume 189, 2013, pp 447-455
3. Le Dortz, N. ; Dept. SSE, Supélec, Gif-Sur-Yvette, France ; Gain, F. ; Zetterberg, P. "WiFi fingerprint indoor positioning system using probability distribution comparison", Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on 25-30 March 2012, ISBN: 978-1-4673-0045-2.

Software Process Development for Healthcare Information Systems Based on Formal Methods

Michael A. Košinár

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
michal.kosinar@vsb.cz

Abstract. Requirements on healthcare software products are becoming more and more complicated and software systems of today are characterized by increasing complexity and size. Therefore, software systems can no longer be developed feasibly without the processes supported by appropriate methods. We propose a method for configuration and modification of agile processes behind healthcare products development based on gathered knowledge and formal modeling. Our approach allows to support and optimize the processes with formal methods of modeling.

1 Introduction

In this paper we introduce software engineering method that focuses on healthcare information systems development domain following best practices given by software processes and formal methods of processes modeling.

We develop a new method and software tools as a part of our research on formal methods, support of software processes modeling, simulation and executing. These tools allow users to model the process with formal methods, simulate it and optimize the whole process model and/or its process execution. This paper focuses on the modeling of the software processes in companies with formal methods to achieve mathematically precise and optimized version of the process.

2 Software Processes and Agile Methods

Business processes represent the core of company behavior. They define activities which companies (their employees) perform to meet their customers' needs. As for the definition of a business process, we quote from Workflow Management Coalition [6]: "Business process is a set of one or more linked procedures or activities which collectively realize a business objective or policy goal, normally within the structure defining functional roles and relationships." A software process is also a kind of business process; it comes with its development and maintenance phases. Our research aims at healthcare and biomedicine information systems development processes. This

can be divided into activities of different kind, as they are defined by the process life-cycle engineering [5].

2.1 Agile Methods

Changes in software products are aspects that are anticipated in agile development, and change is what the modern business software development is all about even in critical systems in a field of healthcare and biomedicine systems.

Implementation of the changes required by software product features or customers' needs might not be easy but agile methods respond to these requirements a give us a solution. The basic paradigm of agile processes is that there should be only the bare minimum of documentation [2] and the production of the actual executable software product as its primary goal. The 'agile' idea is to develop the simplest solution that could possibly work for the current feature [1].

2.2 SCRUM

SCRUM is an agile, lightweight and adaptable process for managing systems development. SCRUM concentrates on how the team members should function in order to produce the required system in a constantly changing business environment. [9]

SCRUM starts with the assumption that software development involves several variables that are likely to change during the project. These include requirements, resources, time frame and technology. SCRUM defines an empirical process control for ensuring that the current status of the project is visible [10].

It is an iterative process: a project developed under SCRUM is divided into sprints, which are the iterations of SCRUM. It is incremental: each sprint produces a working, shippable version of the software with additional functionality [9].

3 Formal Methods

Formal methods are techniques used to describe and model various systems as mathematical models. The big advantage of formal models we utilize in our research is the ability to verify the system's properties and simulate behavior of the system [11]. There are many formal languages that could be used for our needs like finite state automata, Petri nets or Workflow nets, or ontology languages like OWL [12], PROLOG or Transparent Intensional Logic (TIL) [7], [13], [14]. We chose Petri Nets [3] and TIL as a modeling languages as they meet our requirements for its transparency and rich procedural semantics and the ability of Petri Nets to be transformed to TIL [15].

4 Modeling Discipline and Model Notation

We need to define the modeling discipline of processes in order to utilize formal models of software processes in simulation and optimization. The meaning of the term ‘model’ can be understood as a representation of one system – the modeled system – by another system. The modeled system often comes from reality or another artificial complex system and the model is its simplification – abstraction. Three functions of abstraction for modeling a process model are defined in [4, 19, 20]:

- Aggregation: an entity consisting of some other domain entities is modeled as one aggregated entity;
- Classification: a class of entities sharing some common features is identified;
- Generalization: a new class of entities sharing common features is defined by abstracting from those features in which they differ.

However, the software-process development for healthcare and biomedicine domain has some specific features that must be taken into account (see [21]) and it has been characterized as “the most complex endeavor humankind has ever attempted” [8]. Nonetheless, a software process can and should be modeled in a formal way [17].

A knowledge-based approach to software process engineering has been dealt with in [18-20]. The benefits of the knowledge-based approach to modeling and simulation (KBS) in comparison with a Discrete-Event Simulation (DES - [21]) and System Dynamics (SD - [16]) are discussed in [12].

We have 3 basic types of entities that are fundamental for SCRUM process models: Project, Iteration, Task. A Project contains a set of Iterations. Iteration contains a set of Tasks. SCRUM follows an iterative approach to development, using time-boxed cycles. Each release of the system is implemented through a predefined number of time boxed Iterations. Iteration has a start time, duration and, again, contains a set of Tasks. Developers are the main actors/roles of an agile software development process. DES generates experimental data and to facilitate the analysis of all this data, it is conventional to compress the data into a handful of meaningful statistics. In this paragraph are described two distributions from a family of continuous probability distributions. Exponential distribution refers to a statistical distribution used to model the time between independent events that happen at a constant average rate λ . Exponential distribution is used to model waiting time between events.

The Normal (Gaussian) distribution is a continuous probability distribution, and we have used it to generate duration of events (obstacles in development). Proposed simplified agile model entities are (fig. 1)

- Project: This entity represents a software project. It is characterized by Name, a Start Time, a Company and a Team.
- Iteration: SCRUM follows an iterative approach to the development. It has its own identifier (Name), a Start Time, a Duration and contains a set of Stories.

- Story: Represents a single requirements specification of the system. It has its Name, a Type and a Status.
- Task: Each Task has name, a priority, an Estimated Hours, a Remaining Hours, and a Developer.

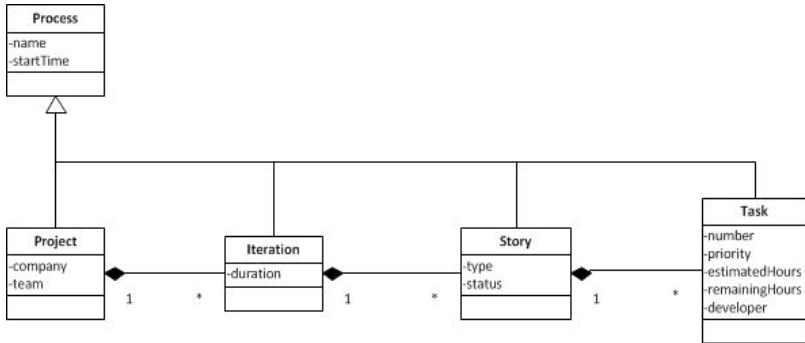


Fig. 1. Extract from Class Diagram of SCRUM Entities

5 Value Added Model

Even though the software products in healthcare and biomedicine are based on similar and rather stable requirements the leading market healthcare and biomedicine companies often utilize SCRUM or other agile methodologies to support and control the development of their products. With more complicated software processes in a field of healthcare information systems development we capture their fundamental and critical parts with formal systems and based on those models combined with knowledge bases we can optimize them and utilize such formal models further in validation, simulation and effort estimation.

Selection of healthcare and biomedicine domain was based on a strong (formal) knowledge behind medicine and hospital fields. With good formal process model and knowledge base behind software products development in software companies on the other side it will be much easier to gather requirements from customers and validate final software products.

6 Conclusion

In this paper we've introduced our approach to modeling of software processes for healthcare information systems development on agile method SCRUM and formal modeling tools. After the brief introduction we've described fundamentals of formal modeling, software process modeling discipline and creation and machine-learning and simulation methods.

Based on our research we introduced the methodology a software tools we have implemented within the development of healthcare information systems for university hospitals and bigger clinics. The realization team of this software product works under the SCRUM and with a good formal model and ontology described in a combination of PROLOG and TIL we could design and build a methodology and support tool for a development process simulation. These simulations help the management to estimate the effort needed for the development of some product parts based on team attributes that are easy to get so they can optimize the process or plan better.

Future work should lead to more complex and precise formal model/ontology of the software process that is executed within the company and a simulation tool that will include more attributes and will be able to work with greater set of external factors that can influence the implementation of the software products. We are also working on our own modeling tool embedded into the simulation application so the tool will be more user friendly to end users.

Acknowledgment, recent publications and awards

Michal Košinár is supported as a *Grand aided student of Municipality of Ostrava, Czech Republic* from 2010.

This research has been supported by the internal grant agency of VSB-TU of Ostrava - SP2013/207 „An utilization of artificial intelligence in knowledge mining from processes and process modeling and mining“

1. Duží, M., Košinár, M., Kožusznik, J., Štolfa, S.: Knowledge-base approach to software-process development based on TIL; In Proceedings of the 22nd European-Japanese Conference on Information Modeling and Knowledge Bases; Editors: Yasushi Kiyoki, Takehiro Tokuda; 2012, Prague, Czech
2. Košinár, M.: Knowledge-base approach to software-process development. In Proceedings of the Wofex 2012, VŠB-TUO, Ostrava, Czech
3. Košinár, M., Štrba, R., Černohorský, J.: Knowledge Modeling and Simulations of Agile Software Process; In Proceedings of the 23rd European-Japanese Conference on Information Modeling and Knowledge Bases; Editors: Yasushi Kiyoki, Takehiro Tokuda; 2013, Nara, Japan
4. Košinár, M.: Knowledge Modeling of Agile Processes in Healthcare Systems Development; In Proceedings of the Trendy v biomedicine 2013, Technická univerzita v Košiciach, Slovakia
5. Košinár, M.: Knowledge Modeling of Agile Processes in Healthcare Systems Development; In ACTA MECHANICA SLOVACA, Slovakia
6. Košinár, M., Štrba, R.: Simulations of Agile Software Processes for Healthcare Information Systems Development Based on Machine Learning Methods; In proceedings of 12th Conference on Programmable Devices and Embedded Systems, PDeS 2013; Velké Karlovice, Czech

References

1. K. Beck. *Extreme Programming Explained: Embrace Change*. Addison-Wesley, 2000. ISBN 0201616416.
2. J. Highsmith and A. Cockburn. *Agile software development: the business of innovation*. Computer, 34(9):120–127, 2001
3. Vondrák. *Methods of Business Modeling*. VŠB-TUO, Czech Republic, Ostrava, 2004.
4. M. Duží, M. Košinár, J. Kožusznik, S. Štolfa (2012) Knowledge-base approach to software process development based on TIL
5. Scacchi W, Mi P (1997) Process Life Cycle Engineering: A Knowledge-Based Approach and Environment. *Intelligent Systems in Accounting, Finance, and Management* 6:83--107-183--107.
6. Workflow Management Coalition: <http://www.wfmc.org/> WfMC Web, 1999
7. Machado EP, Caetano Traina J, Araujo MRB (2000) Classification Abstraction: An Intrinsic Element in Database Systems. Paper presented at the Proceedings of the First International Conference on Advances in Information Systems,
8. Scacchi W (1999) Experience with software process simulation and modeling. *Journal of Systems and Software* 46 (2-3):183-192.
9. K. Schwaber. *Agile Project Management with Scrum*. Microsoft Press, 2003. ISBN 0-7356-1993-X.
10. Scrum. *Control Chaos: Scrum*. Internet, 2006. URL <http://www.controlchaos.com/>. Referenced 22.11.2006.
11. Vondrák, I.: *Neural networks (czech)*. VŠB-TUO, Czech Republic, Ostrava, revision 2009.
12. W3C (2009) OWL 2 Web Ontology Language. <http://www.w3.org/TR/owl2-overview/>.
13. Object Management Group(OMG) (2010) *OMG Unified Modeling Language(OMG UML), Infrastructure, Version 2.3*.
14. Hug C, Front A, Rieu D, Henderson-Sellers B (2009) A method to build information systems engineering process metamodels. *Journal of Systems and Software* 82 (10):1730-1742. doi:10.1016/j.jss.2009.05.020
15. Karkoška, T.: *Vytvoření nástroje pro podporu tvorby ontologií v multi-agentním prostředí*. Diploma thesis, VŠB-TUO, Czech Republic, Ostrava, 2008.
16. Allemang D, Hendler J (2008) *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*.
17. Laurent J-P, Ayl J, Thome F, Ziebelin D (1984) Comparative Evaluation of Three Expert System Development Tools: Kee, Knowledge Craft, Art. *The Knowledge Engineering Review* 1 (04):18-29. doi:doi:10.1017/S0269888900000631
18. Garg PK, Scacchi W (1989) ISHYS: Designing an Intelligent Software Hypertext System. *IEEE Expert: Intelligent Systems and Their Applications* 4 (3):52-63.
19. Mi P, Scacchi W (1990) A Knowledge-Based Environment for Modeling and Simulating Software Engineering Processes. *IEEE Trans on Knowl and Data Eng* 2 (3):283-294. doi:10.1109/69.60792
20. Mi P, Scacchi W (1996) A meta-model for formulating knowledge-based models of software development. *Decis Support Syst* 17 (4):313-330. doi:[http://dx.doi.org/10.1016/0167-9236\(96\)00007-3](http://dx.doi.org/10.1016/0167-9236(96)00007-3)
21. Morgan Kaufmann. Silver GA, Lacy LW, Miller JA (2006) *Ontology based representations of simulation models following the process interaction world view*. Paper presented at the Proceedings of the 38th conference on Winter simulation, Monterey, California

QuickXDB: Native XML Database

Petr Lukáš

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
`petr.lukas@vsb.cz`

Abstract. XML is usually understood as a universal file format for exchanging small data over the Internet. However, XML can also provide an alternative way how to store large databases. Compared to the traditional relational model, XML does not need to be limited by a static schema.

In this paper, we present the current state of our prototype of a native XML DBMS called QuickXDB using state-of-the-art techniques for evaluating XML Queries. We compare the prototype to some other commonly used implementations and show that the prototype can give a better performance.

Keywords: XQuery, XML, query processing, TPQ, DBMS

1 Introduction

Over the years, relational data model has become the most common way of storing large amount of data. However, the relational model is not always the best solution how to organize data. Especially, if we need to have a dynamic structure of a database, we are limited by a static relational schema. XML data model is a useful alternative where the data are naturally organized in a tree¹.

There are many approaches [1, 8, 7, 6] dealing with particular problems around querying XML. This work is focused on putting these pieces together into a real prototype of a robust XML database.

2 Data structures

Since our research is oriented on large databases, we usually cannot load the whole data into the operating memory. Instead, we have to use some efficient persistent structures.

The main data structures of QuickXDB are called *document index* and *partition index*. Both of the indices are B⁺ trees [2], but they differ in what they use as a key (see Figure 1). In the case of the document index, the key is an XML node label (an unique node identifier), while in the case of the partition index it is an XML node name.

¹ The reader is expected to understand the basic terms of XML data model.

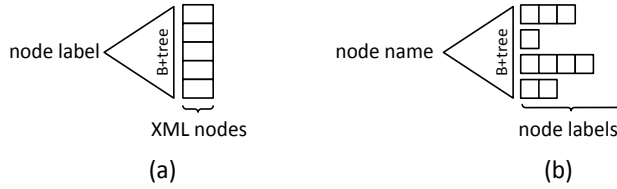


Fig. 1. (a) Document index (b) Partition index

2.1 XSketch synopsis

XSketch synopsis [7] is a subsystem for estimating query result sizes. More concretely we use the f-XSketch variant [3]. Estimating result sizes is crucial for cost-based optimizations described in Section 3.2. XSketch synopsis is a directed labeled graph, where nodes correspond to disjoint sets of XML nodes. The graph is fully loaded into the operating memory to provide immediate response times.

3 XQuery processor

The processor was originally a separated application [5] operating over an in-memory DOM representation of an XML document. It forms the front-end of QuickXDB native XML database. As the name indicates, the processor is capable to treat XQuery queries. We support most of the semantics of XQuery language specified by W3C².

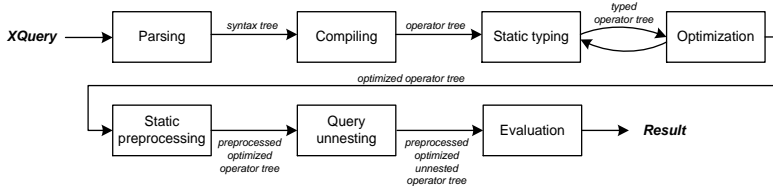


Fig. 2. XQuery processor

XQuery is a representative of declarative program languages, so the queries are not directly interpreted. There is a block schema of the processor in Figure 2. After parsing and compiling a query, the crucial phase is optimization performed together with static typing. Several rewriting rules (e.g., removing unused variables, introducing efficient join algorithms or searching for TPQ (see Section 3.1)) are applied to a tree of operators being constructed. We use an

² <http://www.w3.org/TR/2008/WD-xquery-11-20080711/>

XQuery algebra described in [8], which is a combination of relational algebra and special operators resolving XML structural relationships. The static preprocessing and query unnesting phases produce the final tree of operators (also called a *query plan*), which is subsequently evaluated.

3.1 TPQ rewritings

Twig pattern query (TPQ) is another representation of XML query, expressing core functionality of XQuery. Evaluating single TPQ by a holistic join algorithm (e.g., GTPStack [1]) is usually much more effective than evaluating a complex tree of operators. Identification of TPQ in the original query is not always straightforward. Only a few approaches [6] deal with the problem how to cover possibly the largest fragment of a query plan with a single TPQ. However, they are not fully applicable to our solution, because of different operators and different compilation rules they use.

There is a special operator *TupleTreePattern* with a TPQ as its static attribute. The operator has been already introduced in [6] with a few differences. We do not rewrite the original input query, all the rewriting rules are applied to the tree of operators. We use *TupleTreePattern* and TPQ interchangeably in what follows.

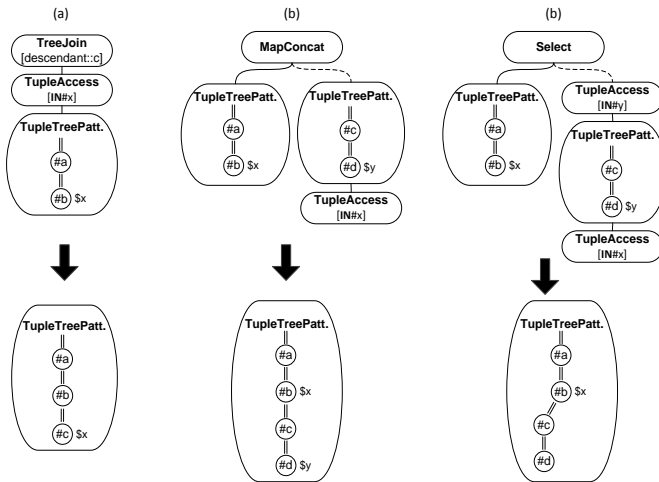


Fig. 3. TPQ rewriting rules

The main principles of our TPQ rewritings are depicted in Figure 3. The first rule (a) merges a single *TreeJoin*³ operator with an instance of *TupleTreePat.*

³ *TreeJoin* represents a single XPath step. This operation is often called as navigational.

tern. The second rule (b) merges two *TupleTreeJoins* under a common *MapConcat* (usually known as Dependent Join) operator. The last rule (c) merges two *TupleTreeJoins* under a common *Selection* operator. Due to space limitations, it is not possible to fully explain the semantics of all operators.

In the ideal case, almost the whole operator tree is rewritten to a single *TupleTreePattern*. However, the TPQ queries can never express the whole semantics of XQuery, so there always have to be other additional operators in the plan.

3.2 Cost-based optimizations

In Section 2, we mentioned that there are two main data structures: document index and partition index. Currently, there are 4 operators accessing data. Two of them use the document index and two of them (including *TupleTreePattern* described in the previous section) use the partition index. It would be better to let the operators choose what data structures will they use.

The operator *TupleTreePattern* is able to combine both methods. We can partition TPQ on parts that will use the partition index and parts that will use the document index. If we want to choose the best partitioning, we need to estimate how long will a query with a given partitioning run.

Cost estimation cannot be done without knowing important characteristics of the data stored in the database. The cost of a query should correspond to the amount of XML nodes read from persistent storage. The XSketch synopsis described in Section 2.1 should be able to give us the desired answer.

4 Experimental evaluation

We have performed an experimental evaluation comparing our processor to other commonly used implementations of XML databases. We have picked up 2 data collections and 10 queries for both of them. Some of the queries were oriented purely on the structure of XML document, i.e. they they resolve only structural relationships between XML nodes and do not access the text content. We call them *structure oriented* queries. The other, which do access the text content are *content oriented* queries. The queries can be found in the technical report [4].

The first data collection, XMark is a 1.2 GB large XML document. The second collection, Treebank, is much smaller – 86 MB, but the structure is highly recursive, so executing queries is complicated in a different way.

4.1 Compared databases

We have tested 5 other XML databases. Oracle database⁴ and Microsoft SQL Server⁵ are mainly used as relational databases. However, they support a fragment of XQuery over XML data. Monet DB⁶ and BaseX⁷ are native XML

⁴ <http://www.oracle.com/cz/products/database>

⁵ <http://www.microsoft.com/sqlserver>

⁶ <http://www.monetdb.org/XQuery>

⁷ <http://www.basex.org>

databases with full support of XQuery. Saxon⁸ is a memory oriented XQuery processor. We have tried various XML indexing options wherever it was possible. Due to lack of space we do not show the results of Microsoft SQL Server and Oracle database systems. We only mention, that they were not able to treat most of the queries.

4.2 Results

The results of the experiment can be found in Table 1. Note that the results shown in this paper slightly differ from the results presented in [4], since we have improved an algorithm for evaluating single-path queries using document index.

XM1 – XM5 and TB1 – TB8 are the structure oriented queries. XM6 – XM10 and TB9 – TB10 are the content oriented queries. All the queries were run 5 times on each database. The values given in the tables are computed as average time without considering the best and the worst time. The best time for each query is highlighted. In many cases a database was not capable to finish executing a query. For example, the XM3 query could not be processed on any database.

As we can see, for the most of the structure oriented queries on both collections, QuickXDB was able to give the best time. The structure oriented queries can be mostly rewritten into a single TPQ, so evaluating TPQs by GTPStack holistic join [1] algorithm seems to be very effective. The content oriented queries run usually slower on our prototype than on MonetDB. This could be caused by the fact that currently there is no data structure indexing content of XML nodes in our prototype.

Table 1. XMark and Treemark experimental results

XMark	XM1	XM2	XM3	XM4	XM5	XM6	XM7	XM8	XM9	XM10
Saxon	0,37	–	–	0,84	0,85	–	2,69	52,94	9,05	–
MonetDB	0,17	–	–	0,52	0,68	0,39	–	0,23	2,08	0,18
BaseX wo/idx	1,13	–	–	1,47	1,84	–	–	252,46	9,04	–
BaseX w/idx	1,02	–	–	1,44	1,95	18,88	–	3,14	9,41	–
QuickXDB	0,07	0,08	–	0,31	0,31	6,6	9,82	6,15	3,5	–
Treemark	TB1	TB2	TB3	TB4	TB5	TB6	TB7	TB8	TB9	TB10
Saxon	0,38	0,49	0,28	0,5	0,76	0,25	1,49	0,37	1,38	0,26
MonetDB	0,32	0,31	0,34	0,45	0,23	0,33	0,38	1,34	0,36	0,33
BaseX wo/idx	1,8	2,54	11,08	2,45	1,88	7,38	8,46	3,06	4,9	1,24
BaseX w/idx	1,81	2,66	11,25	2,24	1,72	7,24	8,31	3,23	5,1	1,26
QuickXDB	0,03	0,02	0,02	0,06	0,15	0,17	0,11	–	0,68	0,22

⁸ <http://saxon.sourceforge.net>

5 Conclusions and future work

We have shown that our QuickXDB, prototype of a native XML database, can give a better performance than some other commonly used XML databases. We have presented the techniques we use to effectively evaluate XML queries and how the techniques work together.

As described in Section 3.2, we are currently introducing cost-based optimizations into our prototype. This should speed up evaluation of a class of queries, since both main data structures – document and partition indices will be used more effectively.

We want to extend the support of XQuery language. Currently, the prototype does not support some significant features like working with XML attributes (the prototype does not distinguish between attributes and actual XML nodes), implicit type conversions or indexing content of XML nodes. The goal is to make the prototype available for other database research communities.

Publications

1. P. Lukáš, R. Bača, and M. Krátký. QuickXDB: A Prototype of a Native XML DBMS. *Dateso 2013*. Pages 36–47. Písek, April 2013. ISBN 978-80-248-2968-5.
2. P. Lukáš. XQuery Algebra. *Information Sciences and Technologies Bulletin of the ACM Slovakia Vol. 4, No. 4*. Bratislava, 2012. ISSN 1338-1237.

Other references

1. R. Bača, M. Krátký, T. Ling, and J. Lu. Optimal and efficient generalized twig pattern processing: a combination of preorder and postorder filterings. *The VLDB Journal*, Pages 1–25, 2012
2. R. Bayer, and E. McCreight. *Organization and Maintenance of Large Ordered Indices*. *Acta Informatica 1*, Pages 173–189, 1972.
3. N. Drukh, N. Polyzotis, Y. Matias. Fractional XSKETCH synopses for XML databases *In Second International XML Database Symposium, XSym 2004*. 2004.
4. P. Lukáš, R. Bača, and M. Krátký. QuickXDB: A Prototype of a Native XML DBMS. *Technical report No. CS/DBRG/2013-001*, 2013. <http://db.cs.vsb.cz/TechnicalReports/CS-DBRG-2013-001.pdf>.
5. P. Lukáš. *XQuery Algebra*. Diploma thesis. Department of Computer Science. VSB-Technical University of Ostrava, 2013.
6. P. Michiels, G. Mihaila, and J. Siméon. Put a Tree Pattern in Your Algebra. In *Proceedings of the 23th International Conference on Data Engineering, ICDE 2007*, Pages 246–255. IEEE, 2007.
7. N. Polyzotis, and M. Garofalakis. XSketch Synopses for XML Data Graphs. *ACM Transactions on Database Systems, Vol. 32, No. 3*. Pages 1014-1063.
8. C. Re, J. Siméon, and M. Fernandez. A Complete and Efficient Algebraic Compiler for XQuery. In *Proceedings of the 22nd International Conference on Data Engineering, ICDE 2006*, IEEE, 2006.

Influence of partiality on procedural isomorphism

Jakub Macek and Lukáš Vích

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jakub.macek1, lukas.vich}@vsb.cz

Abstract. Identity is a basic component of any logic system and ability to recognize that two formulas are identical is key to the process of logical reasoning. In Transparent Intensional Logic we recognize that in addition to symbol-identity there is a more useful form of sameness - procedural isomorphism, sameness of meaning. However the TIL is partial and partiality has great influence on this topic. In this paper we summarize the current results of our research in this influence.

Keywords: TIL, Transparent Intensional Logic, procedural isomorphism, identity, sameness, partiality, alpha-conversion, eta-conversion, beta-conversion, beta-conversion by name, beta-conversion by value

1 Introduction

Identity is a basic component of any logic system and ability to recognize that two formulas are identical is key to the process of logical reasoning. It is obvious that the ability to construct a proof of a conclusion from a set of premises requires us to recognize that the obtained conclusion is the same as the desired one.

While this is straightforward for the first order predicate logic, it is more complicated in cases of higher-order logic systems such as Transparent Intensional Logic (TIL). Since the formula of TIL (called construction) represents an algorithm describing how to obtain a result rather than the result itself, our goal is to recognize whether the algorithm of two constructions is the same.

For this purpose we define several variants of procedural isomorphism. Two constructions are procedurally isomorphic if and only if one can be obtained from the other through a finite series of conversions (such as α -conversion, etc.). The variants of procedural isomorphism differ in the set of permitted conversions. While the variants in themselves represent an interesting topic, in this paper we focus on the elementary conversions.

Since TIL is a type λ -calculus (described in more detail in the next section) the conversions are also based on Church's α , η and β reductions. However another property of TIL, partiality, is in play here and influences which of the conversions can be used and under which conditions. Each of them is described in an individual section in order of increasing complexity.

2 Transparent Intensional Logic

The main concept of TIL is a construction. It is a basic structural element for capturing the meanings of sentences in natural language. Constructions are structured entities and they are defined inductively. But the generated hierarchy has to start with some basic objects - most simple constructions. For these purposes we define two simple constructions: *Variable* and *Trivilisation*. These atomic constructions have no other constituents but themselves. Following definitions are in more detail described in [1] and [2]

- *Variables* are constructions which constructs an objects in depend on valuation (they *v-constructs*, where *v* is a valuation parameter). We can describe how variables (as a constructions) *v-constructs* objects. For every type α we have countably infinite many variables x_1, x_2, \dots . Members of type α can be sorted to infinite many countably infinite sequences $\langle a_{11}, a_{12}, \dots \rangle, \langle a_{21}, a_{22}, \dots \rangle$. Every valuation *v* is a function based on given sequence $\langle a_{i1}, a_{i2}, \dots \rangle$ which assign an element a_{i1} to variable x_1, a_{i2} to variable x_2 etc.
- *Trivilisation* is another atomic construction supplying objects on which molecular constructions are operating. If element of any type α is trivialized, then this trivialization just constructs this element without any change.

Since we have these atomic non-molecular constructions, we can introduce other basic (molecular) constructions. Molecular is because they contains another constituents, not only themselves. These remaining constructions are *Composition*, *Closure*, *Execution* and *Double Execution*.

- *Composition* is an operation of application of function to argument. The purpose is clear - we are doing this for getting a value of function *f* on arguments a_1, a_2, \dots . There are cases, where this procedure may fail. It's strictly there, where the procedure does not provide any output. And this situation occurs when the function *f* on arguments a_1, a_2, \dots is not defined.
- *Closure* is an operation dual to *Composition*. Whereas *Composition* provides that functions can be applied to arguments, *Closure* create whole function by abstracting from values of its arguments.
- *Execution* of construction C is equivalent to C and fails if C is not a construction. Finally, construction of higher orders can be executed twice, that's why *Double Execution*.

Definition 1. *Constructions*

1. Variable *x* is a **construction** which constructs object *A* of appropriate type in depend on valuation *v*. Thus *x v-constructs A*.
2. Let *X* be any object/construction. 0X is a **construction** called *Trivialisation*. It constructs just *X* without any change.

3. Let X, X_1, \dots, X_m be constructions which v -constructs $(\alpha\beta_1, \dots, \beta_m)$ -, β_1 -, \dots, β_m - objects, respectively. $[XX_1\dots X_m]$ is a **construction** called *Composition*. If the function \mathbf{F} v -constructed by X is not defined on the tuple of objects b_1, \dots, b_m v -constructed by X_1, \dots, X_m , then $[XX_1\dots X_m]$ is v -improper, i.e., it does not v -construct anything. Otherwise it v -constructs the value of \mathbf{F} on b_1, \dots, b_m .
4. Let x_1, \dots, x_m be pairwise distinct variables ranging over β_1, \dots, β_m , respectively (β_i, β_j not necessarily distinct types for $i \neq j$) and let X be a construction v -constructing α -objects. Then $[\lambda x_1 \dots x_m X]$ is a **construction** called *Closure*. It v -constructs the following function \mathbf{F} : let $\langle b_1, \dots, b_m \rangle$ be a tuple of β_1 -, \dots, β_m - objects, respectively, and v be a valuation that associates x_i with b_i and is identical with v otherwise. Then \mathbf{F} is undefined on $\langle b_1, \dots, b_m \rangle$ if X is v -improper; otherwise, the value of \mathbf{F} on $\langle b_1, \dots, b_m \rangle$ is what is v -constructed by X .
5. Execution 1X is a **construction** which either v -constructs object v -constructed by a construction X or if X is not a construction or it is v -improper then 1X is v -improper.
6. Double Execution 2X is a **construction**. This construction is v -improper if X is not a construction or if X doesn't v -constructs another construction or v -constructs v -improper construction. Otherwise if X v -constructs construction Y and Y v -constructs object Z then 2X v -constructs Z .

Notice that a composition $[X, X_1, \dots, X_m]$ is v -improper if any of X, X_1, \dots, X_m is v -improper. Let's have some examples of v -proper (resp. v -improper) Compositions.

- i) $[{}^0: {}^0 10 {}^0 2]$ - example of v -proper Composition.
- ii) $[{}^0: {}^0 10 {}^0 x_1]$ - x_1 ranging over τ (real number type), is a v -improper construction for all valuations that associate x_1 with 0.
- iii) $[{}^0: {}^0 x_1 {}^0 0]$ - this Composition is v -improper for all valuations.

It is also notable that Closure is not v -improper for any valuation v . This plays an important role in the following sections.

3 α -conversion

α -conversion is sometimes referred to as *renaming of a variable*. It is the simplest of the three, not only in that the α -reduction is the same as the α -expansion, but it is also quite obvious that it does not change the meaning of the construction.

$$\begin{array}{c} \lambda x [{}^0\text{Prime } x] \\ \Downarrow \\ \lambda y [{}^0\text{Prime } y] \end{array}$$

As the example shows, this is a more or less a cosmetic change and therefore does not alter the procedure of the construction. Also no operation is added or removed, so the partiality does not influence α -conversion in any way.

4 η -conversion

As is illustrated in this following example, η -conversion may, at the first glance, seem as a harmless transformation.

$$\begin{array}{c} \lambda x [{}^0\text{Prime } x] \\ \Downarrow \\ {}^0\text{Prime} \end{array}$$

While this is true for some constructions, it is not for all. The following counterexample was given by Jiří Raclavský at the recent *Epsilon meets lambda* conference. Consider this pair of constructions in case where function F is undefined at the argument A .

$$\begin{array}{c} \lambda x [[{}^0F \ {}^0A] x] \\ \Downarrow \\ [{}^0F \ {}^0A] \end{array}$$

Per definition of Closure the first construction in the pair is definitely not v -improper. Also according to the definition of Composition the second construction may be (and in this case is) v -improper for any valuation v .

Therefore these two constructions are not even equivalent, and so they cannot be procedurally isomorphic (as this is a stronger requirement that entails equivalence). While there is currently no published solution to η -conversion, it has become clear that the core of the problem lies in the change of Closure to Composition (though the core of the solution may be eventually found elsewhere).

We can see that η -conversion is at least allowable for such constructions, where the reduced result may never be v -improper such as with the first case with trivialisation ${}^0\text{Prime}$ or a variable instead of it. Later we refer to this approach as safe η -conversion.

5 β -conversion

The following three constructions are in order: an original (unreduced) construction, a construction after a classic β -reduction (by name) and a construction after a β -reduction by value.

$$\begin{aligned}
& [\lambda x \lambda y [^0 + x y] [^0: ^06 ^00]] \\
& \lambda y [^0 + [^0: ^06 ^00] y] \\
& {}^2[^0\text{Sub } [^0\text{Tr } [^0: ^06 ^00]] \ ^0x \ ^0\lambda y [^0 + x y]]
\end{aligned}$$

The classic β -reduction here resembles the situation described in the previous section. The original Composition that is possibly (in in this case definitely) v -improper is transformed into a Closure which cannot be v -improper. This again leads to non-equivalence.

The β -reduction by value instead describes the proper procedure that is in line with the definition of construction with respect to partiality. First the result of subconstruction $[^0: ^06 ^00]$ is computed. Due to its v -improperness the Composition with Tr function is also v -improper and the same goes for Composition with Sub .

If the inner subconstruction was for example $[^0: ^06 ^02]$, its result would be 3. Per definition of function Tr , the result of that Composition would be a construction such, that it would be a trivialisation of the argument - in this case 03 . This construction is then replaced by function Sub for any occurrence of construction x in the construction $\lambda y [^0 + x y]$. And double execution executes the resulting construction $\lambda y [^0 + ^03 y]$.

Further details on the functions Tr and Sub can be found in [3] and are not key for this paper. The important difference here is however, that the β -reduced construction is again Composition and thus can be v -improper.

Though β -expansion has not been the focus of recent research, an similarly elegant solution is not yet prepared. It should be however noted, that not too complicated approach using Closure is possible.

6 Procedural isomorphism

While the partiality of TIL has brought up some interesting issues in itself, the main goal and result of our research is an extended variant of procedural isomorphism - the almost 1 procedural isomorphism. The numbering refers to Church's equivalences of λ -terms, where '1' referred to a combination of α -, β - and η -conversion.

Definition: Almost 1 procedural isomorphism Let C, D be constructions. Then C and D are procedurally isomorphic, iff there are constructions C_1, C_2, \dots, C_n ($n \geq 1$) such that ${}^0C_1 = {}^0C, {}^0C_n = {}^0D$ and for each pair of constructions C_i, C_{i+1} holds that:

1. C_{i+1} is an α -conversion of C_i ,
2. or C_{i+1} is a safe η -conversion of C_i ,
3. or C_{i+1} is a β -conversion by value of C_i ,

7 Conclusion

In this paper we presented a key part of the problem of procedural isomorphism. While new issues with η -reduction have arisen recently, we are satisfied with the current approach to β -reduction. Though these uncomfortable consequences of partiality may be a convincing argument for non-partial calculus, we are sure that enforcing of total functions is a non-solution, especially in the area of analysis of natural language.

Expanded paper on this topic will be published in the Logica Yearbook.

References

1. Duží, M., Materna, P.: *TIL jako procedurální logika*. Aleph, Bratislava. 2012. ISBN 978-80-89491-08-7
2. Duží, M., Materna, P.: *Constructions*. http://til.phil.muni.cz/text/constructions_duzi_materna.pdf (2000)
3. Duží, M., Jespersen, B., Materna, P.: *Procedural Semantics for Hyperintensional Logic (Foundations and Applications of TIL)*. Springer. 2010. 17. Logic, Epistemology, and the Unity of Science.

Debugging & Profiling of MPI Applications Generated by Kaira

Ondřej Meca and Martin Šurkovský

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{ondrej.meca, martin.surkovsky}@vsb.cz

Abstract. Our research is focused on the simplification of parallel programming for distributed memory systems. Our overall goal is to build a unifying framework for creating, debugging, profiling and verifying parallel applications. The key aspect is a visual model inspired by Colored Petri Nets. In this paper, we will present how to use the visual model for debugging and profiling as well. The presented ideas are integrated into our open source tool Kaira. This paper is based on the joint work with Stanislav Böhm and Marek Běhálek that it will be published at WAPL 2013, Krakow.

1 Introduction

Parallel computers with distributed memory have recently become more and more available. A lot of people can participate in developing software for them, but there are well-known difficulties of parallel programming. Therefore for many non-experts in the area of parallel computing (even if they are experienced sequential programmers), it can be difficult to make their programs run in parallel on a cluster computer. The industrial standard for programming applications in the area of distributed memory systems is *Message Passing Interface* (MPI)¹. It represents a quite low-level interface. There are tools like *Unified Parallel C²* that simplify creating parallel applications, but the complexity of their development lies also in other supportive activities. Therefore, even an experienced sequential programmer can spend a lot of time learning a new set of tools for debugging, profiling, etc.

The overall goal of our research is to reduce complexity in parallel programming. We are developing the tool Kaira³ that is designed as a unified prototyping framework for creating, debugging, profiling and formally verifying parallel applications, where a user can implement and experiment with his/her ideas in a short time, create a real running program and verify its performance and scalability. The key aspect of our tool is the usage of a visual model. In the first

¹ <http://www.mpi-forum.org/>

² <http://upc.lbl.gov/>

³ <http://verif.cs.vsb.cz/kaira>

place, we have chosen the visual model to obtain an easy and clear way how to describe and expose parallel behavior of applications. The other reason is that a distributed state of the application can be shown through such visual model. Based on it, we are able to provide visual simulations where the user can observe a behavior of developed applications. In addition, this feature is available for incomplete applications from an early stage of the development. In context of this paper, the visual model is also useful for debugging and a performance analysis.

On the other hand, we do *not* want to create applications completely through the visual programming. Sequential parts of the developed application are written in the standard programming language (C++) and combined with the visual model that catches *parallel aspects* and *communication*. We want to avoid huge unclear diagrams; therefore, we visually represent only what is considered as “hard” in parallel programming. For more details about Kaira and its visual model see [1, 2].

2 Simulations and tracing of Kaira applications

In this section, we will introduce two crucial features: *simulations* and *tracing* of generated applications. Both can be used for debugging and the latter for profiling.

2.1 Simulations

Besides generating standalone parallel applications from the model, the user can also run the developed application in the simulator. The main task of the simulator is to expose an inner state and it allows for controlling a run of the generated application. The inner state is shown in the form of labels over the original model (see Figure 1). The three types of information are depicted:

- Tokens in place (The state of memory)
- Running transitions (The state of execution)
- Packets transported between nodes (The state of the communication environment)

It completely describes a distributed state of the application. The user can control the behavior of the application by the three basic actions: *start an enabled transition*, *finish a running transition* and *receive a packet from a network*. By executing these three types of actions, the application can be brought to any reachable state.

This approach also gives us the possibility to observe the behavior of the application in a very early state of the development without any additional debugging infrastructure. For example, we can see which data are sent to another process even if there is no implementation of the receiving part.

The user has complete well-formed control of the application in the simulator; therefore, the application can be put into an interesting state (and the user can observe the consequences) even if the application rarely reaches such state.

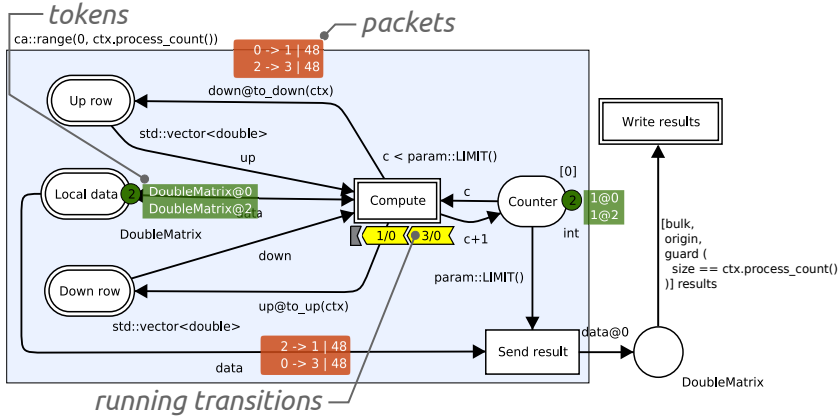


Fig. 1: The model in the simulator

2.2 Tracing

An application developed in Kaira can be generated in the tracing mode, where activities of a run of the application are recorded into a *tracelog*. When the application finishes its run, the tracelog can be loaded back into Kaira and used for the *visual replay* or for statistical summaries. Generally, issues with such post-mortem analysis can be categorized into these basic groups: *selection what to measure*, *instrumentation* and *presentation of results*. Such tracelogs can be useful both for profiling and debugging.

In the case of debugging, we usually want to collect detailed information of the run for the reconstruction of the cause of the problem. In the case of profiling, we want to discover performance issues and therefore need to measure a run with time characteristics as close as possible to real runs of the application. But the measurement itself creates an overhead that devalues the gathered information about performance. Therefore, in both cases, it is important to specify what to store in the tracelog.

In Kaira, the user specifies what is measured in terms of places and transitions. The tracing of transitions enables the recording of information about their execution. The tracing of places enables the recording of information about tokens that go through them. The user can easily control what to measure and it is obvious what information will be gained or lost after switching on or off each setting.

The second task is the *instrumentation*, i.e. putting the measuring code inside the application. In our case, Kaira can automatically place the measuring codes during the process of generation of the parallel application. Parallel and communication parts are generated from the model, therefore we know where are interesting places where to put measuring codes. By this approach, we can obtain a traced version of an application that does not depend on the compiler or computer architecture.

The results are presented to the user in the form of a visual replay or as statistical summaries. In replay, the data stored in the tracelog are shown in the same way as in the simulator, thus as the original model with tokens in places, running transitions and packets on the way and the user can jump to any state in the recorded application. Our tool also provides statistical summaries and standard charts like a normal profiler, and additionally, information is presented using the terms of the model. For example, the utilization of transitions (Figure 2), the numbers of tokens in places, etc.

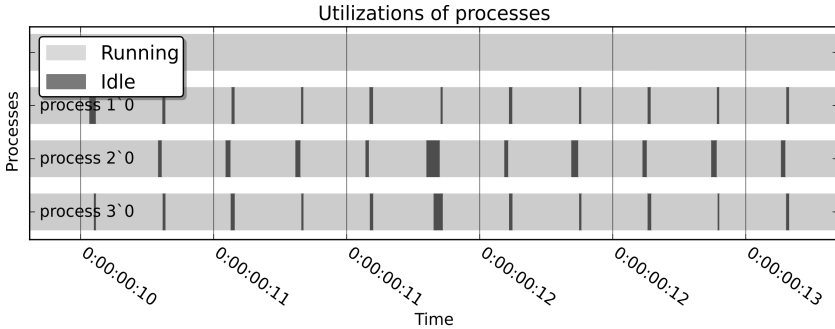


Fig. 2: The example of zoomed chart of process utilizations in the heat flow example with 4 processes.

3 Experiments

This section contains an example where we want to demonstrate features mentioned in Section 2 and compare it with one of the most successful freely available tool *Scalasca* [4]. All programs were executed on a machine with 8 processors AMD Opteron/2500 (32 cores in total) and compiled with Intel Compiler at the optimization level `-O2`.

As an example we use the heat flow problem on a cylinder. We will use a version of this problem where the body is discretized by a grid depicted in Figure 3. The implementation of this problem in Kaira is depicted in Figure 1. The transition *Compute* executes single iteration of the algorithm. It takes a process' part of the grid and two rows, one from neighbor above and one from below. It updates the grid and sends top and bottom rows to neighbors. When the limit of iterations is reached then the results are sent to process 0 where they are written.

On this example we show a comparison between the hand-made solution profiled by Scalasca and the version created and profiled in Kaira. The implementations share the same computation code. It is about 380 LOC (lines of code without comments). The solution in Kaira contains 25 LOC in transitions and

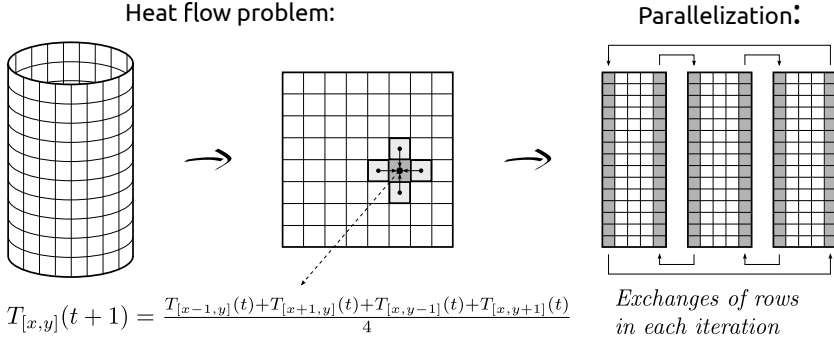


Fig. 3: The heat flow problem on a cylinder and the used method of parallelization

places and 10 LOC for binding the external types. The hand-made solution contains 100 LOC, which are not shared with the solution in Kaira. The following experiments were executed on the instance of the size 6400×32000 and 300 iterations.

All places and transitions are traced except for places *Up row* and *Down row*. The standard function writing token name for the type `std::vector<double>` stores all values from the vector into the tracelog. (6400 doubles for places *Up row* and *Down row*). In our example, we do not need such information, so we can change this writing function to store a smaller amount of data or we can just switch off the tracing (as we have done here). In the case of the hand-made solution profiled by Scalasca, 5 patterns in the filter file was used (23 functions were filtered out). These numbers are small, because of simplicity of the example. Without the filter file, Scalasca produces extremely huge logs (in the order of gigabytes) and it deforms runs of the application, because traced data are very often flushed on the disk.

Table 1. shows the comparison between the solutions generated by Kaira and the handmade solution. In both cases, the measurements were done without writing the resulting matrix into the file. Our problem scales well up to 16 processors, then it reaches limits of used computer.

The measurements show that solution produced by Kaira is comparable to handmade solution and our tracing introduces only a small overhead. For this small number of processors, the measured times are better than the handmade solution profiled by Scalasca. Scalasca is designed for thousands of processors; therefore it is not well suited for this experiment. But our goal was to show that Kaira tracing is comparable (in the scale of tens of processes) with existing mature parallel profilers and Scalasca is a well-established tool in this area.

4 Conclusion

In previous papers, we have been focused on the development of MPI applications by usage of the visual model and visual programming. Our visual language is

Table 1: Measured values for the hand-made solution and Kaira’s solution of the heat flow example

Number of processes	1	2	4	8	16	32
Handmade solution [s]	497.39	249.58	134.39	70.98	57.88	73.38
Handmade solution + Scalasca [s]	3020.89	1525.62	763.63	380.23	193.08	99.33
Kaira solution [s]	443.5	205.57	137.75	72.95	68.04	83.09
Kaira solution with tracing [s]	444.78	229.67	147	72.98	68.14	83.06
Scalasca log size [kB]	128	160	216	336	576	1126
Kaira log size [kB]	40	136	264	520	1126	2150

based on well-known formalism – Coloured Petri Nets. In this paper, we have presented how the same visual model and in fact the same approach was used for debugging and performance analyses.

We introduced a simulator that allows the live introspection into developed programs. This simulator uses the original visual model. Thus the developer is able to inspect the developed application’s behavior using the same visual model that he developed and that he understands.

Also for profiling we use a similar approach and we use the original model. We use it not only to present the obtained data (application’s replay) but also to simplify the measurement specifications. This is crucial for profiling, because when we measure everything, the obtained data are usually useless and setup measurement filters in a standard tool can be hard.

We also demonstrate that presented features can be implemented with a performance that is comparable with existing mature tools. Practical experiments show that a performance of the handmade solution is comparable with the solution generated by Kaira. Measured times differences were up to 20%. The overhead introduced by tracings in Kaira is up to 3%. Our tracelogs are bigger than Scalasca’s tracelogs, but their growths are similar.

This paper is based on the [3] that is joint work with Stanislav Böhm and Marek Běhálék.

References

1. Stanislav Böhm and Marek Běhálék. Generating parallel applications from models based on petri nets. *Advances in Electrical and Electronic Engineering*, 10(1), 2012.
2. Stanislav Böhm and Marek Běhálék. Usage of Petri nets for high performance computing. In *Proceedings of the 1st ACM SIGPLAN workshop on Functional high-performance computing*, FHPC ’12, pages 37–48, New York, NY, USA, 2012. ACM.
3. Stanislav Böhm, Marek Běhálék, Ondřej Meca, and Martin Šurkovský. Visual programming of mpi applications: Debugging and performance analysis. to appear, 2013.
4. Markus Geimer, Felix Wolf, Brian J. N. Wylie, Erika Abraham, Daniel Becker, and Bernd Mohr. The Scalasca performance toolset architecture. *Concurrency and Computation: Practice and Experience*, 22(6):702–719, April 2010.

Background Subtraction using a Layered RC Circuit Model

Karel Mozdřeň, Eduard Sojka, Radovan Fusek, and Milan Šurkala

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{karel.mozdren, eduard.sojka, radovan.fusek, milan.surkala}@vsb.cz

Abstract. The background subtraction is a technique widely used for video analysis, mainly moving object detection for surveillance systems. Such algorithms must be robust, fast and it has to be able to deal with dynamic backgrounds like water surface or moving tree branches. This paper is an extraction of a paper accepted for presentation at 9th International Symposium on Visual Computing ISVC 2013. In this paper, we propose an algorithm for background subtraction based on a model of layered RC circuits. We tested our method on video sequences acquired from level crossing and on commonly used datasets.

1 Introduction

The background subtraction is a common technique for moving object segmentation from video sequences. In many cases, we are not able to predict the size, shape, or color of the objects, that we are trying to detect. In such cases, we are ought to use the background subtraction algorithms. These, instead of training the object detector, model the background and subtract from it the “*moving*” objects in the foreground. The result of background subtraction is a binary image, where the pixels indicating “*moving*” objects are marked with white color and background pixels with black color. Then the connected components algorithm is used to find the individual objects in the segmentation. An example application of such an algorithm is the project Pfinder by Christopher Richard Wren et al. [1]. Their system tracks people and interprets their behavior. The tracking itself is done using the background subtraction algorithm. Generally, PDE based algorithms have the best performance in computer vision applications. Our method is inspired by the model proposed by Pietro Perona and Jitendra Malik [2]. They presented a model, using a grid of resistors and condensers organized in a grid, functioning as an image filter. In this paper we present our modification to this model. We have modified the filter so it can be used for background modeling. The model is extended by excitation contacts used for connection of the input video sequence images with the model. Our paper has the following structure: in Section 2 we describe our modified model, in Section 3 we conduct experiments and compare our method with other commonly used methods, and in Section 4 we conclude our work.

2 Proposed Method

The inspiration to use a layered RC circuit model for background subtraction came from paper by Pietro Perona and Jitendra Malik [2]. In their work, they presented a method for scale-space image filtering based on heterogeneous isotropic diffusion. They also presented a model of diffusion process using electrical components, namely resistors and condensers (RC circuit). The scheme for filter they proposed can be seen in Figure 1. As we can see, the circuit is a grid of

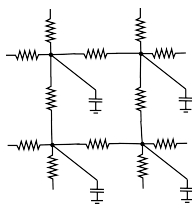


Fig. 1. The image filter model proposed by Pietro Perona and Jitendra Malik

condensers connected to neighboring condensers through the resistors. Each condenser represents one pixel in the image and the voltage is the pixel intensity (or color if we think of the voltage as a vector). If the condenser has lower voltage than its neighboring condenser, then it is being charged by that neighbor (the neighbor is being discharged by it) and vice versa. Charging speed depends on the resistance of the resistors. The greater the resistance is the slower is the charge/discharge rate.

Before we describe modification to the model, we have to understand the background itself and how it differs from foreground. Generally, the background consists of objects that are in most cases stationary and the foreground represents the “moving” objects. Basically, the background model represents the values occurring with higher probability and are not changing that much over time. The foreground objects are represented as values occurring less frequently and are significantly different from the background model.

2.1 Simple RC Model

Our method uses for background modeling a simulation of diffusion using a grid of condensers and resistors as proposed by Pietro Perona and Jitendra Malik [2]. This model is modified in such a way, that each condenser representing one pixel of the background and the voltage over condenser is a representation of pixel intensity. This condenser is connected to excitation voltage representing the input image pixel values through the additional resistor. The one-dimensional example of one block of the original filter and in comparison the modified filter can be seen in Figure 2. This way the video sequence values are filtered

and the voltage u_c over the condenser C represents the filtered value (modeled background), which is an equivalent of the mean μ value used in TGBS. The standard deviation parameter $u_{c\sigma}$ can be modeled similarly using the absolute difference between input excitation voltages (values) and modeled background u_c as an excitation voltage for u_σ circuit. The differential equation describing the background update has the following form

$$\frac{\partial u_{c,x,y}}{\partial t} = \frac{1}{CR} (u_{c,x-1,y} + u_{c,x+1,y} + u_{c,x,y-1} + u_{c,x,y+1} - 4u_{c,x,y}) + \frac{1}{CR_e} (u_{e,x,y} - u_{c,x,y}) , \tag{1}$$

where the u_c is the background model value (voltage over condenser), x and y represent the position in the circuit grid, C is the condenser capacity, R is the resistance value of the resistors connecting the condenser C with the neighboring condensers, and the resistor R_e connects the excitation voltage u_e (new value) to the background model condensers. The magnitude of the excitation resistance

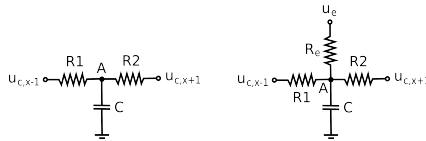


Fig. 2. The scheme of one block of the original (left) and modified filter (right)

R_e regulates the speed of background model adaptation. If the resistance is too small, the adaptation speed is fast, and if it is big, the adaptation speed is slow. Fast adaptation leads to imprint of slowly moving objects into the background model. On the other hand, if the adaptation is too slow, some parts of the background image might be outdated.

This model is most similar to the TGBS method. The main difference between these two methods is in filtering. The TGBS filters the data only in time domain, because it filters the values for each input image pixel separately, but our method considers the background model as a whole, where all the pixels are connected to theirs nearest neighbors, which allows filtering also between background model pixels.

2.2 Layered RC Model

The simple layered RC model is already able to distinguish between foreground and background, but it still needs to be modified for adaptation to the dynamic backgrounds. Firstly, this was solved by C. Stauffer and W.E.L. Grimson [3] in MoGBS. They developed a method, where the background is not modeled only

by one Gaussian, but by a mixture of K Gaussians. There, each of the Gaussians adapts to one kind of the backgrounds emerging in the video sequence.

Inspired by this method, we introduce the layered RC circuit model (LR-CBS). We added to our circuit (model) additional layers, that are used to represent multiple backgrounds, the *DEMUX* which is a demultiplex connecting the input voltages u_e (input values) to specific layer, which is selected by selector S . Each layer is also connected by the inter-layer resistor R_L providing inter-layer filtration. This helps mainly in initialization step, if the backgrounds are set randomly, and also when one background disappears. Layer, that does not represent any background moves towards to the nearest active background layer, where it helps to represent its background. The modified background modeler scheme can be seen in Figure 3. The adaptation of this model is driven by selector S . When

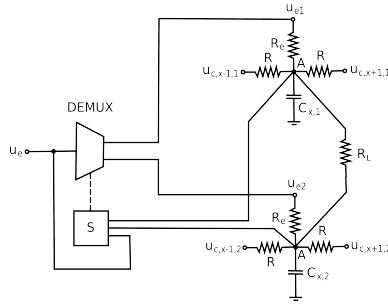


Fig. 3. The scheme of one one-dimensional block representing the layered RC model

the new excitation value u_e emerges, the selector S connects it using the demultiplex *DEMUX* to the layer, where the absolute difference between new value u_e and mean value $u_{c,l}$ is minimal. The other layers excitation values $u_{e,l}$ are leveled to their corresponding voltages $u_{c,l}$ (no excitation). This is performed for each block in the model and then the following difference equation is used for the update

$$\begin{aligned}
 u_{c,x,y,l}^{(t+1)} = & u_{c,x,y,l}^{(t)} + \frac{dt}{RC} (u_{c,x-1,y,l}^{(t)} + u_{c,x+1,y,l}^{(t)} + u_{c,x,y-1,l}^{(t)} \\
 & + u_{c,x,y+1,l}^{(t)} - 4I_{c,x,y,l}^{(t)}) + \frac{dt}{R_e C} (u_{e,x,y,l}^{(t)} - u_{c,x,y,l}^{(t)}) \\
 & + \frac{dt}{R_L C} (u_{c,x,y,l-1}^{(t)} + u_{c,x,y,l+1}^{(t)} - 2u_{c,x,y,l}^{(t)}), \quad (2)
 \end{aligned}$$

where $u_{c,x,y,l}^{(t)}$ is the value representing the background for pixels at position x, y in the layer l at time t . The dt is a time difference constant, and R_L is the resistance of the resistor connecting individual layers. Similarly, the u_σ is modeled. The pixel is marked as foreground, if the input value is not within the distance of 2.5σ from the most similar layer value.

3 Experiments

In this section we experiment with the complete method and we compare it to other commonly used methods. The experiments were conducted on real video sequences captured from IP cameras installed on a level crossing, that we use for tests of obstacle detection and obstacle behavior monitoring. Furthermore, we use standard datasets, often used for testing and comparison of background subtraction algorithms.

The first experiment that we have conducted was a test of the abilities of the proposed method under different conditions like dynamic backgrounds and changing illumination. Resulting segmentations in Figure 4 clearly show that our method is able to adapt under many difficult conditions. The tests were

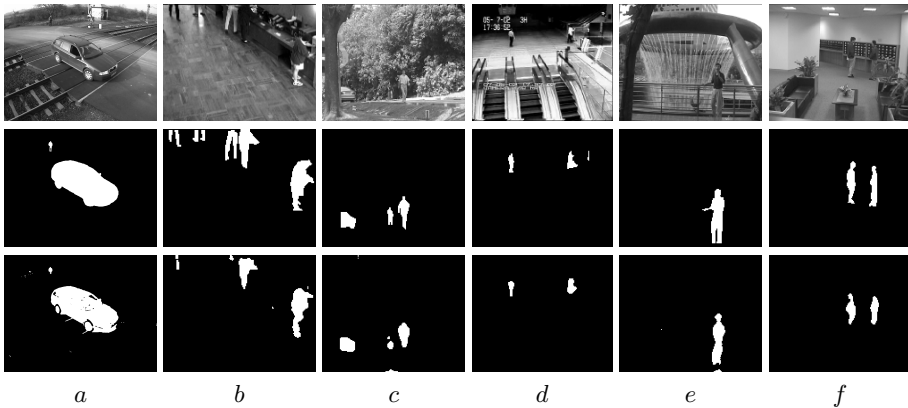


Fig. 4. Testing images their foreground segmentations and ground truths. First row) input images; Second row) ground truth; Third row) Our segmentations, a) Level Crossing; b) Bootstrap; c) Campus; d) Escalator; e) Fountain; f) Lobby

conducted using the following configuration: $R = 2 M\Omega$, $C = 1 \mu F$, $dt = 0.001$, $R_L = 2 M\Omega$, $R_{min} = 20 k\Omega$, $R_{max} = 200 k\Omega$, $\lambda = 5.0$, $K = 5$.

Our second experiment is quantitative. We compared our method with other commonly used methods. In this test, the level crossing dataset has been used. This dataset consists of high resolution images, and therefore, moving object detection can be measured more accurately. As a quantitative measure the previously presented Matthews Correlation (Phi) Coefficient (MCC) [4] is used. The results can be seen in Table 1. There, you can see that our method performs better than other commonly used methods. The layered version LRCBS is in addition able to deal with dynamic backgrounds. It is clear from the experiments that it outperforms the MoGBS.

Table 1. Algorithms comparison: LRCBS: layered RC; MoGBS: mixture of gaussians; TDBS: temporal difference; TGBS: temporal gaussian; TMBS: temporal median

	TP	TN	FP	FN	Phi
LRCBS	681225	20385616	103824	62999	0.89
MoGBS [3]	489913	20329902	295136	118713	0.7
TDBS [5]	461445	19816336	323604	632279	0.48
TGBS [1]	630959	18605493	154090	1843122	0.41
TMBS [6]	535682	17712377	249367	2736238	0.29

4 Conclusion

We have developed a novel algorithm for background subtraction using a layered RC circuit for background modeling. We have shown in the experiments that our algorithm is able to adapt to many difficult conditions like strong illumination changes, casting shadows, and dynamic backgrounds. Furthermore, our method performs better than other commonly used methods for background subtraction.

Acknowledgements This work was supported by the SGS in VSB Technical University of Ostrava, Czech Republic, under the grant No. SP2013/185, and Ministry of Industry and Trade of the Czech Republic - project TIP No. FR-TII/027 .

References

1. Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.: Pfunder: real-time tracking of the human body. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **19** (1997) 780 –785
2. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **12** (1990) 629 –639
3. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. In: *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. Volume 2. (1999) 2 vol. (xxiii+637+663)*
4. Powers, D.M.W.: Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. Technical Report SIE-07-001, School of Informatics and Engineering, Flinders University, Adelaide, Australia (2007)
5. Hironobu, A.L., Lipton, A.J., Fujiyoshi, H., Patil, R.S.: Moving target classification and tracking from real-time video. (1998) 8–14
6. Lo, B., Velastin, S.: Automatic congestion detection system for underground platforms. In: *Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on. (2001) 158 –161*

An Application of Process Mining to Real Business Processes

Jakub Stolfa

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jakub.stolfa@vsb.cz

Abstract. One of the key issues of all companies is to try to optimize the cost of everyday operations. Company management would like to know how are operations or processes executed and if there are any deviations that waste much money. There are many ways how to support the process enactment, but at the end when there are some data from the process usage the analysis of the efficiency is needed. One of the ways how to analyze the process and effectively analyze the process data is to use process mining methods. In this paper, we present the usage of process mining methods to real invoicing process and show the possible impact of the results to the process or organizational improvement. Paper also describes research in the area of process estimation and process modeling that is connected with the process mining.

Keywords: process mining, case study, SAP, process improvement, process estimation

1 Introduction

Companies enact most of their key business processes with the support of information systems. Enacted processes are represented by workflows, partly managed by the information system, partly managed by users' decisions and activities. It is not easy to understand whether specific process runs efficiently, because usually many various activities are processed in parallel and process definition allows plenty of process enactment variations. Our task was to analyze specific process with request to suggest steps for its simplification, curtailment and enact the process cheaper.

Our paper is organized as follows: Section 2 introduces the state of the art; Section 3 depicts the experiment that we have performed: describes the data that we have obtained; shows the usage of the process mining methods and explains obtained results and its interpretation in real business; concluding Section 4 provides a summary, discusses the planned future research and explain connection to our other research.

2 State of the Art

In last decades, systems started to be more and more process oriented [1]. The shift from the data oriented systems to the process oriented systems brought the companies tools to control and check the enactment of the processes and resources that are involved.

Business process definitions are sometimes quite complex and allow many variations. All of these variations are then implemented to supportive systems. If you want to follow some business process in a system, you have many decisions and process is sometimes lost in variations. Modeling and simulations can help you to adjust the

process, find weaknesses and bottleneck during the design phase of the process. If you want to analyze the real usage of the system, recognize its weaknesses, bottlenecks or strongness on the real data, you have to know how the process was followed in reality. Process mining is an approach that is used for the analysis of real enactment of the processes. Process mining uses logs of real process enactments to analyze the process itself [4].

2.1 Related work – case studies

There is a lot of papers that describe new ways or improvements of methods, techniques and algorithms used in the process mining. Surprisingly only several papers are focused on the case studies, that mean in fact practical use of the process mining. Detailed overview of recent case studies up to 2011 can be found in [5]. Authors describe 11 case studies in several domains, mainly in public services.

The range of process mining case studies and variations of process mining types and methods shows the wide range of process mining applicability to answer different questions in different domains.

3 Case study

This cases study reacts to a company request to analyze business process of the invoice verification in SAP system with aim to identify context in which the process is not effective and provide a suggestion to the process improvement. The analyzed company runs SAP system in five countries (with five different jurisdictions) and process approx. 30 000 supplier invoices per year.

3.1 Context

Examined business process of the invoice verification is implemented in SAP ERP and SAP DMS, user activities are controlled by SAP business workflow. Users participate in the invoice verification workflow in several different roles (creator, accountant – completion, approver, and accountant – decision and posting).

Generally, it is process where the accountant should create the invoice, verify it, then send to the approvers and finally when he gets it back he does invoice posting. The case study is about verification of the idea about the process, find deviations and in case of it do retrospective view to the particular instance of invoice verification.

3.2 Data obtaining and data structure

The main task during data preparation was to collect and adjust the application logs to the final structure (see Table 1) needed for the process mining analysis. Data was collected from the workflow system log containing most information about activities.

Table 1. Data structure of the log

IDOBJ	ID of the invoice	mandatory
IDACTIVITY	ID of the activity	mandatory
DATESTART	Date when the activity started	optional
TIMESTART	Time when the activity started	optional
DATEEND	Date when the activity ended	mandatory
TIMEEND	Time when the activity ended	mandatory
ROLE	Role of the invoice processor (creator, accountant,	mandatory

	approver)	
USER	User account of the invoice processor (real user name)	optional
TRANSACTION	System transaction run by the user during processing the activity	mandatory
DATA	Detailed information about context of the invoice – specifically ORGID, INVOICETYPE, ...	optional

We have loaded the log with DATEEND between 1. 1. 2012 - 30. 6. 2012, totally we loaded **37 991 records** for adjusting.

3.3 Preprocessing of data

The main task of the data preprocessing is to map records from the application log to the parts of the process in the process mining.

We recognized following event classes:

- Invoice *Creation* – this event in the real business process means completion of the invoice (scanned originals, completion of all important fields) or recompletion of the existing invoice.
- Invoice *Verification* – this event verifies the formal correctness of the current invoice (VAT, accounts, ...)
- Invoice *Approval* – this event approves the invoice (responsible persons approve the factual correctness of the invoice and approve or reject the invoice, or send the workflow back to the Verification/Creation)
- Invoice *Posting* – alternative decision of accountant (largely the final posting of the invoice and change of the responsible approver)

The analysis was performed on **36 711 records**, or respectively events.

3.4 Analysis of process model

Process model is depicted in the figure Fig. 1. We have reconstructed process model from the log by BPMN Analysis (using Casual Net Miner) [11]. Process starts by artificial event Start and then continues by event *Creation*. Next one is *Verification*. These two events *Creation* and *Verification* can be done repeatedly. Next, there is a gateway where the *Approval* event can be skipped. *Approval* event can be done repeatedly too. Last events are *Posting* and artificial event *End*.

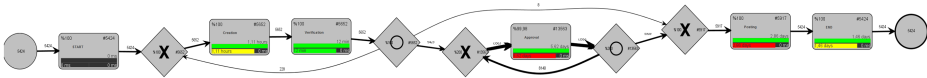


Fig. 1. Process model

The process model shows even the frequency of the transitions between events. It is depicted by numbers next to arcs or by thickness of arcs. When the arc is bolder, then the path is more frequented. We can see the high frequency of regular repeating of *Approval* event. It means that invoice is *Approved* more than once quite often.

Other interesting finding is that *Creation* and *Verification* events are done always together sequentially.

We have discovered the most frequented paths, respectively all the used paths. Number of all used paths (patterns) is 76. The Table 2 shows summary of the most

frequented and most interesting paths that were discovered. Detailed analysis of these most frequent paths is provided in the next section.

Table 2. Interesting paths in the process model

Path	Case(s)	Occurrence (relative)	Number of events
Path 1 – 1 st most frequented	2436	44.912%	7
Path 2 – least time consuming path	1	0.018%	10
Path 3 – without <i>Approval</i> event	1	0.018%	5

3.5 Analysis of interesting paths

This section describes the analysis of most interesting paths that we chose from all paths of the process model.

Path 1 – 1st most frequented path

This path is used by most process enactments (cases) of the log. It is almost the half of all cases. The process model of this path is depicted on the Fig. 2. Process starts by *start* artificial event, continues by *Creation* event, *Verification* event. Next, there are **two** repetitions of *Approval* event and process ends by *Posting* event followed by the *end* event.

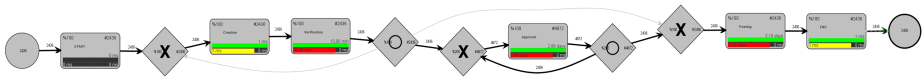


Fig. 2. Process model - Path 1

Path 2 – least time consuming path

This path is least time consuming path. Same as like most time consuming it includes only one case. The path is depicted in the Fig. 3.

Process of this path, or case, is following:

1. *Start* artificial event
2. Three times – *Creation* and *Verification* event
3. *Approval* event
4. *Posting* event
5. *End* artificial event

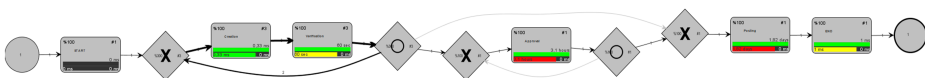


Fig. 3. Process model - Path 2

Path 3 – without Approval event

We chose this path for detailed analysis because it do not contains *Approval* event. This is some kind of deviation we are looking for. The path is depicted in the Fig. 4.

Process of this path, or case, is following:

1. *Start* artificial event
2. *Creation* and *Verification* event
3. *Posting* event
4. *End* artificial event

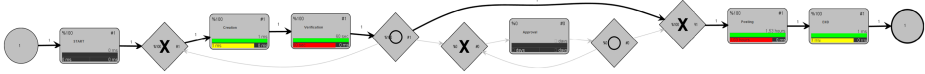


Fig. 4. Process model - Path 3

3.6 Results and retrospect to real business

We have analyzed specific paths and we have found following results:

- **Path 1:** The invoices from this class were found as invoices connected to the purchase order/contract and with proper receipt without differences (amount, price). Typically there are invoices for the investments, strategic raw material and overhead invoices with two persons approving the invoice. This proves that the purchasing system is well defined and settled.
- **Path 2:** there was not found any workflow for specific invoices, there are notes at these invoices with link to the documents with confirmation (meeting notes) – it means the invoice was confirmed by competent people, but out of the workflow system. Such path must be analyzed separately.
- **Path 3:** repetition of the “posting” step always means that the first approver of the invoice refused the responsibility for this invoice (step “not me”). Such decision has mostly reason in assignment of competencies, sometimes in mistakes in invoice classification or other typing error. Monitoring of this path is suggested for the future.

4 Conclusion and future work

There are some conclusions that we have found about the invoice verification process using process mining techniques.

1. Very specific cases were found in Path 2 – they do not have any frequent pattern and from this reason are not interesting for conclusion.
2. Path 3 provides the information about the model of competencies and it is appropriate to monitor this path.

This case study opened several tasks for future work. First of all, we have not paid attention to the resources. In future work, we would like to focus also on users, their interaction during the process and construction of their social network

We have been concentrated to most frequent paths with the aim to find reserves and improving themes – in the future, we would like to focus also on deviations that can mean some specific problem.

Our study proved that even the basic process mining methods used to analyze the enactment of the processes can be very useful for the real businesses. The processes performed by humans and supported by different systems have often wide range of variability, thus it is not always possible to predict and control the real usage of the process even with the sophisticated system. Thus we have to use information from the real enactments to see what really happened. We can reconstruct the process by the process mining tools; find deviations and other many useful pieces of information by still grooving bunch of process mining methods. This analysis can be then used to improve the process, supporting system or the organizational behavior, to estimate process behavior. Our intention for the future is to study the usage of the process min-

ing methods, find their best value for the business and try to present new methods that can answer the questions that arise from the real business needs.

Our research is also focused on estimation of process and its parameters and on process modeling. This can be also used in the connection with the process mining. There is a lot of task to resolve so we will continue with our research in this way. Papers where we describe our research are: [12-17].

References

- [1] M.Dumas, W.M.P. van der Aalst, A.H.M. ter Hofstede, *Process Aware Information Systems: Bridging People and Software Through Process Technology*, Wiley- Interscience, 2005
- [2] D. Grigori, F. Casati, M. Castellanos, U. Dayal, M. Sayal, M.-C. Shan, *Business process intelligence, Computers in Industry* 3 (2004) 321–343.
- [3] M.Weske, W.M.P. van der Aalst, H.M.W.E. Verbeek, *Advances in business process management, Data & Knowledge Engineering* 50 (1) (2004) 1–8.
- [4] W.M.P. van der Aalst, A.H.M. ter Hofstede, M. Weske, *Business process management: a survey, Business Process Management* (2003) 1–12.
- [5] Jochen De Weerd, Annelies Schupp, An Vanderloock, Bart Baesens, *Process Mining for the multifaceted analysis of business processes—A case study in a financial services organization, Computers in Industry, Volume 64, Issue 1, January 2013, Pages 57-67, ISSN 0166-3615, 10.1016/j.compind.2012.09.010.*
- [6] Lee, D., & Bae, H. (2013). Analysis framework using process mining for block movement process in shipyards. *ICIC Express Letters*, 7(6), 1913-1917.
- [7] Accorsi, R., & Stocker, T. (2012). On the exploitation of process mining for security audits: The conformance checking case. Paper presented at the Proceedings of the ACM Symposium on Applied Computing, 1709-1716.
- [8] Li, M. (2012). Process mining in knowledge maintenance a case study. *Advances in Information Sciences and Service Sciences*, 4(9), 293-301.
- [9] Van Dongen, B.F., De Medeiros, A.K.A., Verbeek, H.M.W., Weijters, A.J.M.M. & Van Der Aalst, W.M.P. 2005, "The ProM framework: A new era in process mining tool support", *Lecture Notes in Computer Science*, pp. 444.
- [10] Van Der Aalst, W.M.P., Van Dongen, B.F., Günther, C.W., Mans, R.S., Alves De Medeiros, A.K., Rozinat, A., Song, M., Verbeek, H.M.W. & Weijters, A.J.M.M. 2007, "Process mining with ProM", *Belgian/Netherlands Artificial Intelligence Conference*, pp. 453.
- [11] Van Der Aalst, W., Adriansyah, A. & Van Dongen, B. 2011, *Causal nets: A modeling language tailored towards process discovery*.
- [12] Štolfa, J., Štolfa, S., Koběřský, O., Kopka, M., Kožuszník, J., Snášel, V.: *Methodology for Estimating Working Time Effort of the Software Project*, In: 2012 Databases, Texts, Specifications, and Objects (DATESO), pages 25-37, 2012
- [13] Štolfa, J., Koběřský, O., Kopka, M., Krömer, P., Štolfa, S., Kožuszník, J., Snášel, V.: *Value estimation of the use case parameters using SOM and fuzzy rules*, In: The International ACM Conference of Emergent Digital EcoSystems (MEDES), 2012
- [14] Jakub Štolfa, Svatopluk Štolfa, Jan Kožuszník, Tereza Moudrá: *Business Process Formal Modeling in Graphical Ontology Tool - Functional View*, In: The International ACM Conference of Emergent Digital EcoSystems (MEDES), 2012
- [15] Jakub Štolfa, Ondřej Koběřský, Pavel Krömer, Svatopluk Štolfa, Martin Kopka and Václav Snášel: *Comparison of Fuzzy Rules and SVM Approach to the Value Estimation of the Use Case Parameters*, IFSA World Congress NAFIPS Annual Meeting, 2013
- [16] Svatopluk Štolfa, Jakub Štolfa, Pavel Kromer, Ondrej Kobersky, Martin Kopka and Vaclav Snasel: *Fuzzy Rules and SVM Approach to the Estimation of Use Case Parameters*, The 4th International Conference on Innovations in Bio-Inspired Computing and Applications, Ostrava 2013
- [17] Martin Kopka, Miloš Kudělka, Jakub Štolfa, Ondřej Koběřský and Václav Snášel: *Extraction and Analysis Social Networks from Process Data*, 5th International Conference on Computational Aspects of Social Networks (CASoN), 2013

Hierarchical Layered Mean-Shift Methods

Milan Šurkala, Karel Mozdřeň, and Radovan Fusek

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{milan.surkala.st, karel.mozdren, radovan.fusek.st}@vsb.cz

Abstract. Many image processing tasks exist. Segmentation is one of the tasks we deal with, especially with the mean-shift method. Our goal is to improve its speed and reduce the over-segmentation problem that occurs with small spatial bandwidths. We propose new mean-shift method family called Hierarchical Layered Mean Shift. It uses hierarchical preprocessing and stacking hierarchical segmentation outputs together to minimise over-segmentation. Significantly better speed should be achieved and suppression of the over-segmentation of the layered versions should be preserved or even improved.

Keywords: mean shift, segmentation, acceleration, hierarchy

1 Introduction

Segmentation is the image processing tasks that should find significant objects in digital images. It is still constantly developing and the goal is to improve not only the accuracy and the segmentation quality but also the speed of algorithms. We are focused on the method that is called Mean Shift. It was released in 1975 [6], but it started to be developed more 20 years later, in 1995 [2]. Very important papers about this method are, for example, [3], [5], and [4].

The main idea of the mean-shift method is in iterative motion of data points to the position of their highest density. In digital images, these data points are image pixels. For each pixel, a kernel density estimate is computed and it is shifted according to this estimate. It is repeatedly computed until the point converges to an attractor, the place of the highest density of pixels. In general Mean Shift (MS) [6], we need two datasets. The first one is the original dataset that is used to compute the density estimate of the data points and the second one holds the shifted values (the actual data points for which the density estimate is computed). If we use Blurring MS [1], only one dataset is needed. The source dataset is replaced by the computed values after each iteration. The BMS method also has the smaller number of iterations per data point.

The hierarchical approaches [7] showed to be a very fast way to accelerate the algorithms with a very small influence on the segmentation quality. They use more stages of the algorithm with different bandwidths and they use the output from the first stage as the input for the following one. A small bandwidth in the first stage ensures fast computation as well as the smaller input in the next one.

The layered approaches, on the other hand, run several MS computation and stack them together. If the pair of pixels belongs to the same segment in more resultant segmentation (threshold has to be set, for example, in two segmentations of the three processed), they are merged together.

We present Hierarchical Layered Mean Shift (HLxMS) methods that use the hierarchical approach and stacking of segmentations to reduce the over-segmentation problem. The “x” stand for unspecified mean-shift method, for example, hierarchical layered mean-shift method using the Blurring MS approach is denoted as HLBMS. The version using the general MS, is denoted basically as HLMS. Difference between hierarchical and layered approach are visible in figures Fig. 1 and Fig. 2 respectively.

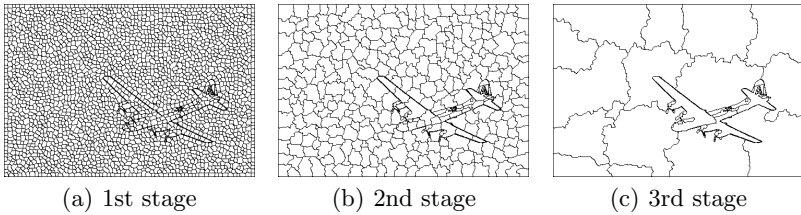


Fig. 1. Stages of the HBMS method.

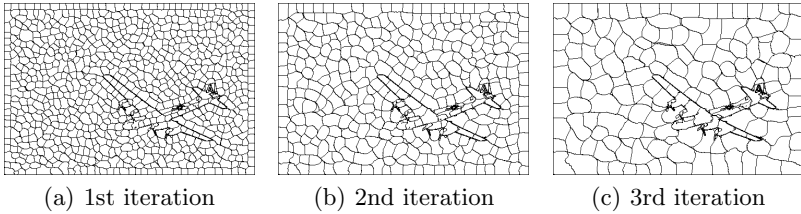


Fig. 2. Iterations of the LBMS method.

2 Mean Shift Methods

Let $X = \{x_i\}_{i=1}^n \subset R^d$ be a dataset of n points in the d -dimensional space, the *kernel density estimator* for BMS method is given by the following equation

$$p(x) = \frac{1}{n\sigma^d} \sum_{i=1}^n K\left(\frac{x - x_i}{\sigma}\right), \quad (1)$$

where the first fraction is a normalisation constant and σ is the bandwidth. It sets the diameter of the searching window (kernel function $K(x)$). We use two

types of bandwidths in digital images. The spatial bandwidth σ_s is limiting the neighbourhood of the processed pixel in x and y axis. The range bandwidth σ_r indicates the maximum possible luminance difference between the processed pixel and the pixels in its neighbourhood. MS can use broad kernels (for example, the *Gaussian*) that cover all the data points and σ parameters change only the shape of kernel as all pixels are involved in computation. If we use truncated kernels (for example, the *Epanechnikov, uniform*), the bandwidth parameters limit the size of kernel. In our algorithm, we use only truncated kernels, because it is based on using the small kernels that improve the speed of the algorithm. The Epanechnikov kernel is given by the equation

$$K(x) = \begin{cases} 1 - x^2, & \text{if } \|x\| \leq 1 \\ 0, & \text{otherwise} \end{cases} . \quad (2)$$

The *mean-shift vector* that is iteratively needed for each pixel in the processed image, is given by

$$m_{\sigma,k}(x) = \frac{\sum_{i=1}^n x_i k\left(\left\|\frac{x-x_i}{\sigma}\right\|^2\right)}{\sum_{i=1}^n k\left(\left\|\frac{x-x_i}{\sigma}\right\|^2\right)} - x, \quad (3)$$

where the function $k(x)$ is a derivative of the kernel $K(x)$. This equation indicates the difference between the former position of the processed pixel x (on the right-hand side of the equation) and a new position of the processed pixel x . In each iteration, the point is moved to the new computed position until the movement is zero or close to zero. Each iteration consists of moving all the points according to their mean-shift vectors, than the output of this iteration is converted to the input for the next one.

3 Hierarchical Layered Mean Shift

Our new method called Hierarchical Layered Mean Shift (HLxMS in general) combines the hierarchical and layered approach. The hierarchical approach divides the segmentation of the image into several stages. The first stage use very small spatial bandwidth and, therefore, it is carried out very fast. It creates a large number of small segments. We consider the segment as one data point with the weight proportional to the number of points it contains. This output is used as the input for the next stage with a larger spatial bandwidth. Because this input is smaller due to preprocessing in the first stage, the next stage can be carried out very fast too even if the larger spatial bandwidth is used (the dataset is smaller). HBMS results can be seen in Fig. 1.

The layered approach uses a larger number of stages on the same dataset. We carry out computation with several small, but different bandwidths. Each result has the same boundaries around the significant objects but it has variously shaped segments in the flat areas of the image. If we stack the segmentation results, the most significant boundaries would be clearly seen (Fig. 3). The *merging*

algorithm is straightforward. If two random pixels are, for example, twice in the same segment in two different segmentations of three processed segmentations, they are assigned to one bigger segment.

We can carry out much more segmentations and we need to set the threshold t lower than this number to denote the number of segmentations, where two random pixels need to be in the same segment in order to assign them to the same final segment. We have to check all the pairs of pixels that are in the distance equal or slightly higher than the maximal used spatial bandwidth (we do not need to check all the pairs of pixels in the processed image). It greatly reduces the computational cost of the algorithm with no influence to the segmentation result.

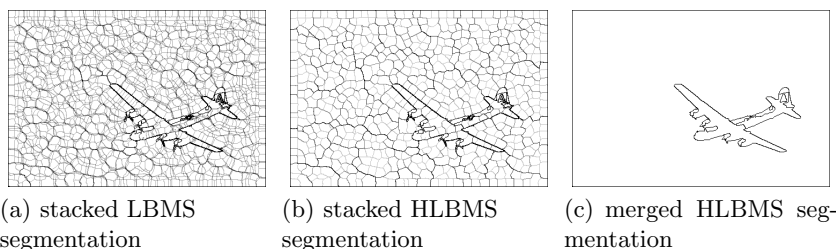


Fig. 3. Stacked LBMS and HLBMS (hierarchical LBMS) image and the result after merging the HLBMS segments.

HLxMS combines both approaches together. It carries out one very fast initial segmentation with a very small bandwidth. Its output is used as the input for two or more layered segmentations. General layered approach uses the original image as an input, whereas the hierarchical layered approach uses the first pre-processing stage for creating its input as in the hierarchical versions of mean-shift methods. Simply, the hierarchical approach makes one over-segmented image and the layered segmentation post-processing will decide which segments should be grouped and which boundaries should be preserved. Of course, more hierarchical stages can be executed in order to minimize the computational time if we process very large images.

4 Experiments

We present the experiments carried out with Blurring MS (BMS) and Evolving MS (EMS) [8]. General MS is not presented as it is not very suitable for such stacking. We show original single-layered algorithms (BMS, EMS), hierarchical versions (HBMS, HEMS), layered versions (LBMS, LEMS) and new hierarchical layered versions (HLBMS, HLEMS).

The speed comparison can be seen in Table 1. It is obvious that new hierarchical layered approach slightly improved the speed of algorithms and almost completely reduces the over-segmentation problem.

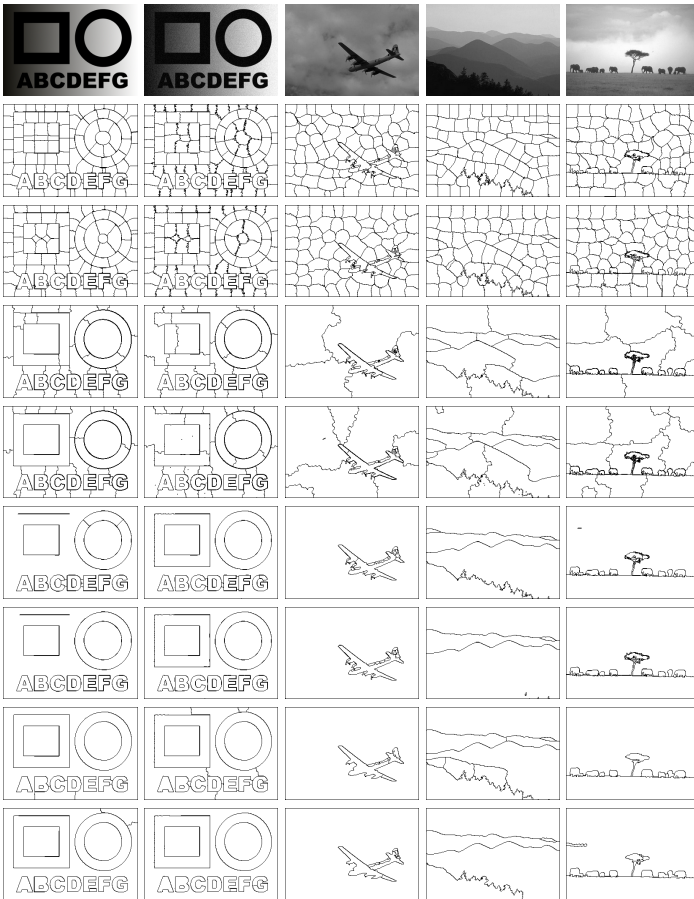


Fig. 4. Rows 1: the original image; 2/3: BMS/EMS (spatial bandwidth $\sigma_s = 20$); 4/5: HBMS/HEMS ($\sigma_s = 4/16/64$); 6/7: LBMS/LEMS ($\sigma_s = 4$, multiplier of the bandwidth $mul = 1.4$); 8/9: HLBMS/HLEMS ($\sigma_s = 3.5$, $mul = 1.35$).

Table 1. The comparison of speed depending on algorithm

	synthetic	syn. noise	airplane	hills	savana
BMS	31.8 s	37.5 s	34.7 s	40.2 s	34.7 s
HBMS	8.7 s	10.1 s	9.6 s	8.9 s	8.2 s
LBMS	42.2 s	43.4 s	46.1 s	44.3 s	41.1 s
HLBMS	6.4 s	5.6 s	6.1 s	6.4 s	6.1 s
EMS	812.2 s	603.8 s	168.3 s	198.9 s	160.9 s
HEMS	29.7 s	33.8 s	18.8 s	17.3 s	20.4 s
LEMS	561.5 s	440.2 s	149.2 s	156.2 s	153.2 s
HLEMS	20.3 s	20.7 s	16.8 s	17.0 s	17.8 s

5 Conclusion

We have shown that the layered approaches can almost eliminate the oversegmentation problem in a reasonable time. The general layered versions are sensitive to parameter settings. Especially BMS and EMS algorithms are very useful in their layered versions. Hierarchical layered versions that we have presented in this paper do not suffer from the need of proper setting of parameters and they work very well with general MS. It outperforms the hierarchical mean-shift methods in both areas of the speed and the segmentation quality.

References

1. Carreira-Perpiñán, M.: Fast nonparametric clustering with Gaussian blurring mean-shift. In: Proceedings of the 23rd international conference on Machine learning. pp. 153–160. ICML '06, ACM, New York, NY, USA (2006)
2. Cheng, Y.: Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 790–799 (1995)
3. Comaniciu, D., Meer, P.: Mean shift analysis and applications. *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* 2, 1197–1203 vol.2 (1999)
4. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(5), 603–619 (May 2002)
5. Comaniciu, D., Ramesh, V., Meer, P.: The variable bandwidth mean shift and data-driven scale selection. *Computer Vision, IEEE International Conference on* 1, 438 (2001)
6. Fukunaga, K., Hostetler, L.: The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory* 21(1), 32–40 (January 1975)
7. Šurkala, M., Mozdreň, K., Fusek, R., Sojka, E.: Hierarchical blurring mean-shift. In: Proceedings of the 13th international conference on Advanced concepts for intelligent vision systems. pp. 228–238. ACIVS'11, Springer-Verlag, Berlin, Heidelberg (2011), <http://dl.acm.org/citation.cfm?id=2034246.2034270>
8. Zhao, Q., Yang, Z., Tao, H., Liu, W.: Evolving mean shift with adaptive bandwidth: A fast and noise robust approach. In: ACCV (1)'09. pp. 258–268 (2009)

Introduction to financial markets trading

Jiri Urbasek

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
urb279@vsb.cz

Abstract. This article presents a brief introduction into current state of financial markets trading. It describes markets dynamics that causes price changes of financial assets and states most popular algorithms used nowadays to predict this changes. Finally article provides a summary of popular trading software including their capabilities.

Keywords: market dynamics, algorithms, trading, biologically inspired, trading software

1 Introduction

Trading financial markets is becoming very popular and with today's technologies available almost anyone can start trade markets at almost no initial costs. In past it was much more difficult to become trader. One had to call broker by phone, create account, and then call broker to initiate every transaction buy and sell commands. Broker then would try to pass those commands to person present at floor of trading house and execute the command physically.

Nowadays, it is much easier. It is possible to establish everything online, transfer money from bank account and become live trading.

At this article we describe basic fundamentals of how financial markets behave, how can we predict market behavior and what software and platforms can be used to accomplish trades on various financial markets.

2 Financial markets behavior

Financial markets uses principles of auction to satisfy supply and demand of participants. The two main auction principles incorporated in markets are called English auction and Dutch flower auction. Typical form of English auction used is called "open outcry". It means that on seller offers goods to multiple buyers that make a bid on price they are willing to pay for the goods. Highest bid wins. The Dutch auction works in the opposite way. Seller puts an item at high price and continually decreases until buyers accept the current price.

There is two main ways of analyzing financial market behavior. Fundamental and technical analysis. In **fundamental analysis** we decide value of specific

market product from events that relate to it. Eg. political, geographical aspects, climatic conditions or any other events that can influence the value. Investors who invest in accordance with fundamental analysis must watch news permanently and individually decide on impact any information can have on their trade. Another approach of predicting market behavior is **technical analysis**. This type of analysis uses only mathematical and statistical instruments and does not pay attention to any events happening to financial instrument in real world. Computers helped greatly with evolvement of technical analysis. Typically predictions are made by applying mathematical and statistical methods to previous market behavior. Predictions are made especially while inspecting three main market attributes - price, time and volume of traded contracts at a specific time.

In this paper we will focus only on technical analysis instruments.

2.1 Efficient-market hypothesis

Efficient-market hypothesis was developed in early 1960s by Eugene Fama. It states that there exists three forms of market efficiency depending on how much is market determined by public information:

- in **weak-form efficiency** market behavior prices are determined by all publicly available information thus it is not possible to predict future prices from historical values. Fundamental analysis is suitable for this type of market efficiency and can provide good trading results
- **semi-strong-form efficiency** means that market reflects all public information so quickly, that it is very difficult by investors to constantly beat market values
- **strong-form efficiency** assumes that prices fully reflect all public and private information and that it is impossible to achieve excessive trading returns consistently.

2.2 Financial market theories

Many theories have been created about market behavior. There exist many investors who advocate *Random walk Theory* saying that market price randomly fluctuates around it's real value and change in value cannot be predicted from historical information.

Other theories use waves and cycles to explain price changes. *Elliot wave principle* claims that prices creates 5 waves in direction of main trend (up or down) and 3 waves in opposite direction (price correction).

There is many different market behavior theories. In any of them in general we can say that price changes in respect of following paradigms:

- **trading market**: this type of market is characterized by price fluctuation in a relatively low range and is most commonly seen. Price rises up until it

reaches price level called resistance level. From that point price goes down until it reaches support level – lowest price. Then it continues back up and repeat the cycle.

- **trending market:** this is market with clearly defined up or down trend. Up trend is called bull market. Down trend is called bear market. Bull market is usually slow and covers greater time period whereas bear market is often very quick.
- **changing market:** this is period at which market is going to change from trading to trending or vice versa.

3 Technical analysis instruments

First technical analysis attempts were made by Charles Dow in 19th century when he developed his Dow Theory. This theory uses two market indices - *Dow Jones Industrial Average* (DJIA) and *Dow Jones Transportation Average* (DJTA) to analyze market behavior.

Dow theory is described by six main principles:

1. Market price is determined by all public and private information. This include economic and psychological factors and other behavior of all traders.
2. There is three trends that define market. Primary trend, lasting for months or years, it contains secondary trend lasting for weeks, that contains minor trends lasting less then few weeks.
3. Trend goes through three major phases. Accumulation – well informed investors start follow the trend. Public participation – majority of investors realize trend and start to trade it. Distribution – positive economic news attract a mass of investors which results in change of trend direction.
4. Both DJIA and DJTA must confirm each other to clearly indicate signal of an coming trend.
5. Trend is confirmed by trading volume. In bull trend volume increases, in bearish trend volume decreases.
6. A trend ends at the moment when signal of opposite trend is confirmed.

3.1 Trading indicators

Trading indicator is type of algorithm used to analyze financial markets. There is literally hundreds of different indicators used nowadays. We can divide them into several types according of what they try to measure.

Trend indicators Trend indicators are used to identify trend in a market. Examples are Moving averages or Moving Average Convergence-Divergence (MACD).

Momentum indicators These indicators can be used to detect overbought or oversold positions and to determine possible start of new trend. Examples are Commodity Channel Index (CCI), Relative Strength Index (RSI) or Stochastic Oscillator.

Volume indicators Volume indicators take advantage of information about volume of traded goods. This information can tell a lot about future possible move of a market. Examples are Chaikin Money Flow and Money Flow Index.

Volatility indicators These indicators try to show behavior of traders and advise entry signals for new trades. Fo example Bollinger Bands.

4 Technical aspects of trading

There exists many of exchange trading software systems. In this section we will cover the most frequent ones and compare their price and trading techniques here.

4.1 Exchange of trade information

Several network protocols were developed to allow data transfers between commodity exchange and others who want to use data for their purposes. There is some widely used protocols as well as some others used only by specified exchanges that developed their algorithms for their own purpose. Commonly used protocols include FIX(Financial Information eXchange), FAST (FIX Adapted for STreaming), ITCH , HSVF (High Speed Vendor Feed), STAMP (Securities Trading Access Messaging Protocol) and others.

FIX FIX is a textual protocol used to transfer market and trades information. Official documentation[4] defines FIX as:

”The Financial Information eXchange protocol was invented by Salomon Brothers and Fidelity Investments in the early 1990s to address the problem of communicating securities orders and their executions in the wholesale banking industry. Since the early 90s FIX has been expanded, through 7 protocol releases to cover all asset classes and extend its reach fromtrade initiation to trade confirmation. FIX is today the most widely adopted protocol in the wholesale markets and the defacto standard for all institutional trading globally.”

FAST As electronic trading increased in volume it became evident that FIX protocol is not efficient enough and too verbose. This led to unwanted delays in communication. From this reason FAST protocol was developed. In a nutshell FAST acts as thin layer above FIX protocol that removes data redundancies resulting into significantly smaller message size.

ITCH Protocol with ultra low latency.

HSVF This protocol is designed and developed by Montreal Exchange (owned by TMX Group). It is part of Sola Trading Solutions.

4.2 Trading platforms and software

TradeStation Most recent service description is located at the official webpage <http://www.tradestation.com>.

Trade station is company based in USA that provides one of the most popular trading platforms. It offers access to most of the used markets such as stocks and ETFs, options, futures, forex, mutual funds and bonds.

It has its own programming language called *EasyLanguage* to allow programming of automatic trading strategies.

It offers various pricing options. Stock, ETFs and options can be priced either per share or flat-fee commission from each trade. Futures and forex are priced as commission from each trade.

Oanda USA based company that is specialized at providing forex trading platform. Service is located at <http://www.oanda.com>.

Oanda supplies real-time forex data that can be used in trading softwares to perform trades. It is possible to create two types of accounts. Demo account is free and allows to practice trades on real data with unlimited virtual funds for unlimited time. It behaves exactly as real trades would behave but operates only with virtual money. This type of account is convenient for testing trading strategies. Live account is identical to demo account but operates with real money. Oanda earns money as a little commission from each trade.

Instaforex Is Asia based forex broker company. Their service is located at <http://www.instaforex.com>.

Instaforex works very similarly as Oanda. It offers two types of accounts – demo and live with only difference, that demo account operates with virtual money.

NinjaTrader Is options and futures trading platform that has it's own software called NinjaTrader. All information is located at <http://www.tradersplatform.com>.

It offers only 30-day trial free service. After that point it is paid as a commission from each trade.

Plus500 London based company that delivers trading facilities on commodities, forex, shares, ETFs, CFDs and indices. Their service is located at webpage <http://www.plus500.com>.

As similar to other platforms, Plus500 offers two types of accounts. Demo account that has no limitations and operates virtual money. And live account. Plus500 platform includes software for Windows, web browser, iOS and Android.

MetaTrader MetaTrader is not broker platform. It is software, that is able to connect to platform providers and perform trades on their platform via API. Currently most recent version is version 5 and is free to download at

<http://www.metatrader5.com>. MetaTrader cooperates with various brokers such as Oanda. It is only for Windows computers.

5 Conclusion

At this article we have briefly described the ways of how can financial markets be traded nowadays. We stated some basic theories that try to explain why financial markets behave the way they behave. These theories cannot be applied to all possible cases and thus one cannot rely fully on them. At next part we have described how trading works at physical way. Which protocols and systems are used to transfer trade information from trader to exchange house. Predicting of future prices can be also based on historical data. There exist lots of algorithms called market indicators that use historical data to predict future market moves. These are briefly described in last part of the article. Together with list of most commonly used broker platforms and software to perform market trades.

References

1. Patricak Gabrielsson. *Evolving biologically inspired algorithms*
2. George Drasňnar. *Hazardní hry: Úvod do spekulace s cennými papíry*
3. David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*
4. Kevin Houstoun. <http://fixprotocol.org/documents/2801/FIX> Adapted for STreaming - FAST Protocol.pdf, *FIX Adapted for STreaming - FASTsm Protocol Technical Overview*
5. Burton G. Malkiel. *A Random Walk Down Wall Street*
6. TradeStation – Online Trading — Trade Stocks, Options, Futures & Forex Online — Trading Software. <http://www.tradestation.com>
7. OANDA – Forex Trading and Exchange Rates Services. <http://www.oanda.com>
8. instaforex – Instant Forex Trading. <http://www.instaforex.com>
9. NinjaTrader – Online commodity Trading. <http://www.tradersplatform.com>
10. Plus500 – World’s Trading Machine. <http://www.plus500.com/>
11. MetaTrader 5 Trading software. <http://www.metatrader5.com>

An analysis of an object's parameters by means of image processing

Rafał Gasz

Opole University of Technology,
Faculty of Electrical Engineering, Automatic Control and Computer Science,
ul. Prószkowska 76, 45-758 Opole, Poland
r.gasz@doktorant.po.edu.pl

Abstract. High quality assurance is an important element of any production process. Therefore, an important factor is a comprehensive quality examination carried out quickly, thoroughly and in a cost effective manner. This article presents a method of examining the quality of cast elements by means of image processing techniques.

1 Introduction

The most effective quality control is carried out in a dynamic manner. It makes it possible to take into consideration all processes occurring in a particular object's production by identifying and anticipating quality-related problems.

An effective and quick identification of defects requires an examination after the completion of each stage of production. In this respect, there are two possible approaches - supplied components can be examined before they are used subsequently in the production process or the production process can include an auto-control system operated by a person manufacturing or processing components.

From the technical point of view, the objective of quality control is to check, e.g. by measuring or benchmarking, a product's one or several characteristic properties and compare the acquired results to the required ones for the purpose of confirming their compliance with the design. This task is usually performed by the dedicated personnel whose operating activities do not include production workers' duties. Products which do not comply with specifications are rejected or returned for corrections [1].

2 The methods used in assessing the quality

The simplest method of product quality control is the precise measurement and assessment of a product's condition by the qualified technical personnel. Examinations can be carried out by means of simple measuring devices such as meters, slide callipers, micrometer gauges and comparative control models. An assessment of a

product's condition includes also a visual evaluation of workmanship or the placement of a manufactured element in a properly prepared mould or master die, which allows one to verify the correctness of the production process. Such methods are simple and cheap. Their main disadvantage is a long execution time and consequently - low efficiency. In such methods here is also a probability of major mistakes in comparison to automated methods of evaluating a product's quality.

Another type of measuring methods comprises the use of coordinate measuring machines. Measurements taken by means of coordinate measuring machines are characterized by a very high degree of accuracy and objectivity. A considerable advantage of this measurement technique is the possibility of taking measurements of objects with complicated shapes which cannot be measured by means of basic workshop measuring tools such as slide callipers, microscopes or dial gauges. With respect to casts which are to be subjected to further machining, the use of coordinate measuring machines is inadequate in view of their costs and the required measuring precision at a level of $(50 \div 100) \mu\text{m}$. A measuring robot constitutes a solution offering a comparable functionality. In comparison to coordinate measuring machines, measuring robots are characterized by a higher speed of the arm up to 0.5 m/s. Their acceleration is also a few times higher up to 4 m/s^2 . The operating conditions of a measuring robot located in a cabin can correspond to those of factory production halls [2].

The optical shape measurement methods ensure a high resolution as well as a variable sensitivity and scope of measuring the surfaces of three-dimensional objects. Depending on technique, they can be used to measure distances as well as a three-dimensional micro- and macrostructure of a tested object. They can be divided into two basic groups: point and field methods. In the point method, information on an object's shape is acquired from a particular set of points isolated from its surface (similarly to coordinate measuring machines). On the other hand, in the field method, information on an object's shape is acquired by means of pictures of the whole object under examination recorded by means of a camera.

The point methods are related to active and passive triangulation techniques. Their basic disadvantages include a limited number of points which can be analyzed in the case of fully automated data processing and a long time of data processing.

The field methods constitute a much larger group of the optical shape measurement methods. Their major advantage is the fact that they provide information on an object's shape simultaneously from the whole field of vision. This generation's methods are based on the analysis of the intensity of deformed fringe images projected onto the surface of a tested object. Projected fringe images can be created by means of various techniques, e.g. the projection of a grid image or the creation of interference fringes from a coherent light source. Changes in height and deformation of the surface of the tested object cause interferences in the fringe pattern of a known shape, which is subsequently compared to the original.

Two structures in the form of linear rasters imposed on each other cause the appearance of an image of dark and light bands of moire. They are dependent on the frequency of spatial structures, the ratio of the widths of dark and light lines and the grids' mutual positioning.

A raster image placed in the projection branch is projected onto the object. In the other branch of the system there is an identical optical system which maps the fringes

from the object onto the reference raster, the same as the raster of the projection branch. The rasters' planes are perpendicular to the optical axes, which, in turn, are parallel to each other. The rasters are shifted transversely and, in order to illuminate the largest possible surface of the tested object, a high degree of precision is necessary in the adjustment of the optical systems. An analysis of fringe images can be treated as a process of a complete recreation of a measured value encoded in such images, constituting sinusoidal intensity fluctuations [3]. The optical shape measurement technique based on the moiré effect is presented in Figure 1.

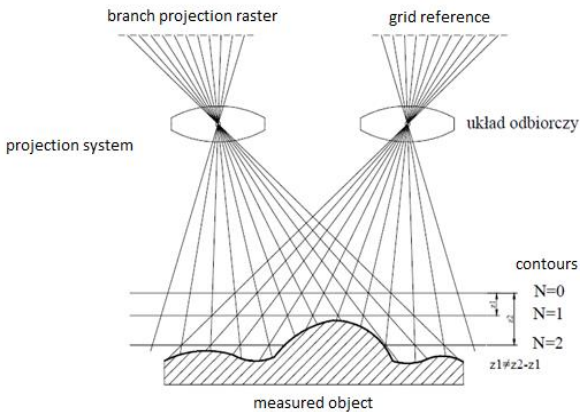


Fig. 1. The phenomenon of projection moiré

Apart from the aforementioned optical methods based on geometrical relations, calibrated photographs can be used in some measurements. Product examination can be performed indirectly by comparing photographs of a model to photographs of a particular detail, or by detecting particular shapes. This method is presented in more detail in the following chapter. The application of the optical method in the examination of the quality of casts could be a quick and relatively cheap method as compared to, for example, the use of a measuring robot. In view of a short duration of measurement, all casts could undergo examination.

3 Analysis of the object

As an object is to be examined in a short period of time, about 2–3 seconds per point and with accuracy at the level of 100 μm , an examination method has been developed whose measurement determination procedures are based on the analysis of images acquired from a high-resolution camera.

The method requires the maintenance of constant illumination conditions and a repeated state of an object's surface. In this method, the compliance of photographs of a particular detail with a model can be defined with respect to selected fragments, e.g. holes. The method is simple and quick in application because it requires the taking of just one photograph. Photographs can be taken by means of a camera with a standard lens. As it has already been mentioned, no measurements are taken in this method, only comparisons are made. Because of the ascending perspective, the same details

located at various distances are visible on photographs as details with different dimensions. Figure 2a includes deformations which appear if a standard lens is used. The use of a telecentric lens (Figure 2b) eliminates such deformations, which can cause serious problems with computer-based image analysis.

In this method, image deformations are not important because the use of the same measuring system causes the occurrence of deformations on both the model and the inspected object. For example, only holes located on the lens's axis are circles. As the distance from the axis increases, circles change into ovals.

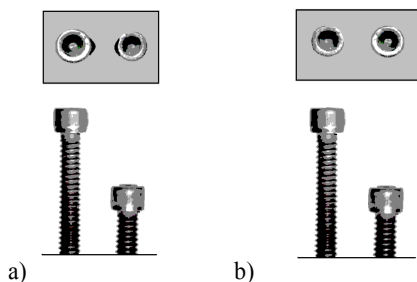


Fig. 2. Photos a hex screws of different lengths of the lens: a) standard b) telecentric

Practical experience shows that it is difficult to achieve the method's good effectiveness with respect to the complete consistency of images. The method's effectiveness can be improved by using transformed images, e.g. edge detection (Figure 3) so that only selected elements (e.g. holes' edges), and not the texture of an object's surface, are compared.

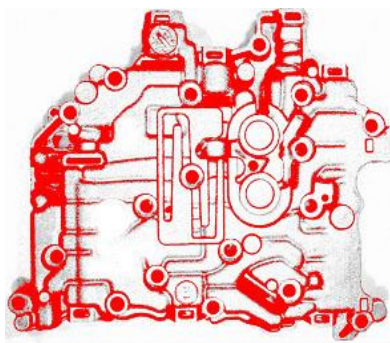


Fig. 3. Image after processing, prepared to compare [4]

The measurement of the diameters of the holes on the cast's side turned out to be a difficult problem. The diameters of these holes decrease stepwise. Because of the holes' considerable depth and the consequent half-shadow, the taking of measurements with top illumination does not give the expected results. Thanks to cracks through which light can reach certain sections of the hole, this problem can be solved by illuminating the cast from the bottom (Figure 4). Such illumination makes it possible to take measurements in a few sections [4].

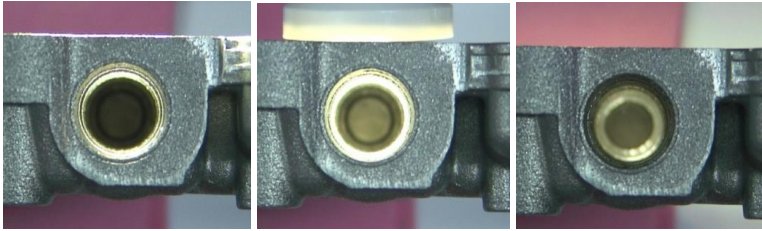


Fig. 4. Lighting of the lateral holes at different depths hole [4]

Such measurements required the use of a multipoint controllable illuminator and an additional holder for the inspected element. The most difficult problem to solve was the measurement of the height of the cast's selected points in relation to the plane of reference. Assuming that the cast's support occurs under three points (determining the plane of reference), the measurement of the cast's upper edge (if the cast's thickness in a given place is known) can confirm, e.g. the presence of twists. We took test measurements of the height by means of the triangulation method, illuminating a selected point at a known angle (60°) with a laser beam (Figure 5).

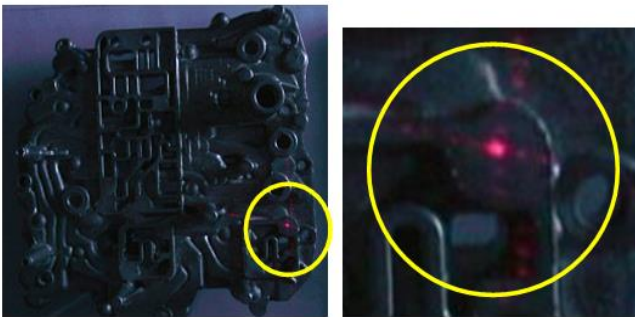


Fig. 5. Lighting selected the laser beam [4]

In order to record the image, the object needs to be darkened because of low laser light scattering on the cast's surface in the case of lateral illumination. In the second solution, the selected place is illuminated with two cones of light (an apex angle of more than 30°), with shifted foci necessary for the determination of the sign of deviation. In the event that the plane is at the determined position, two circles will be visible: one with a minimum diameter of about 1 mm, and the other – 2 mm (Figure 6). The indirect measurement of the height consists in the measurement of the diameters of both circles [5].



Fig. 6. Lighting two points to measure the height of the object [5]

An evaluation of an object's parameters should include optical defects which worsen an image's parameters, thus impairing the accuracy of taken measurements. In order to eliminate occurring deformations, the photographs underwent the deformation correction process.

4 Conclusions

Image processing has numerous applications in various fields. Due to low costs of implementing necessary software, we can acquire a handy tool for a comprehensive evaluation of cast elements. Currently the most popular method of evaluating the condition of such elements is the use of appropriate moulds or master dies. This method is cheap, but relatively slow. The method described in the paper is quick and accurate. More accurate measurement results can be acquired by means of coordinate measuring machines; however, evaluation costs are disproportionate to the task.

An important aspect of the discussed method is the application of adequate diagnostics equipment. The solution proposed by the authors for evaluation of holes' distribution and shapes can constitute a quick and accurate tool for evaluation of the workmanship of cast elements.

References

1. Hamrol A., Mantura W.: Zarządzanie jakością. Teoria i praktyka. Wydawnictwo Naukowe PWN, Warszawa, 2002
2. Ratajczyk E.: Współrzędnościowa technika pomiarowa. Oficyna Wydawnicza Politechniki Warszawskiej, 2005
3. Sojecki A.: Optyka. WSiP, Warszawa, 1977
4. Zator S. Gasz R.: Kontrola jakości odlewów z użyciem analizy obrazu, [w:] Knosala R. (red.): Innowacje w Zarządzaniu i Inżynierii Produkcji, Oficyna Wydawnicza PTZP, pp.1023-1033
5. Gasz R., Zator S.: Measurements of casting components using image processing methods, Electrodynamic and Mechatronic Systems editors: Bronisław Tomczuk, Andrzej Waindok, Jan Zimon, Dawid Wajnert, Zawiercie 2013, Oficyna Wydawnicza Politechniki Opolskiej, pp.49-50



Rafal Gasz is a recipient of a Ph.D. scholarship under a project funded by the European Social Fund

EUROPEAN UNION
EUROPEAN
SOCIAL FUND



The Use of Absolute Vibrations for On-line Monitoring of Industrial Electric Motors

Marek Klimowski

Opole University of Technology,
Faculty of Electrical Engineering, Automatic Control and Informatics,
Institute of Industrial Electrical Engineering and Diagnostics,
ul. Prószkowska 76, 45-758 Opole, Poland
m.klimowski@estudent.po.edu.pl

Abstract. The possibilities for using absolute vibrations for on-line monitoring of industrial electric motors have been presented in this article. An engine applied in the cement industry and made on sliding bearings served as an example. In the article there have been included the following descriptions: of the measurement technique, the measurement system, the applied sensors and the vibration signal transducers. The rules have been described which should be followed while selecting the spot where the control equipment is to be installed as well as the assembly method. Possible sources of interferences, which can cause serious failures of the performed measurement of the vibration signal, have been indicated. The diagnostic possibilities in a continuous operation of material electric machines have been presented in the study after performing the vibroacoustic and spectral analysis of the occurring absolute vibrations.

1 Introduction

Ensuring continuous and safe operation of electric machines applied in industry is a basic aim of each production company. It is necessary to promptly recognize changes of dynamics of the supervised machines, the level of wear of material parts, the type and level of damage. Maintenance services make use of various on-line systems of diagnostic testing whose measurements enable reducing the number of failures and serious damage to critical electric motors. Vibration diagnostics based on the measurement and assessment of relative and absolute vibrations of the tested electric engine is one of the best monitoring methods. In numerous cases vibration measurements carried out on non-rotating parts of machines are enough to characterise properly their current motor conditions with respect to parameters of proper work. They are used, first of all, for the current operational supervision and acceptance tests. General conditions and measurement procedures, assessments of vibrations of non-rotating parts of machines, as well as general valuation criteria are all included in the PN-ISO 10816-1: 1998 standard Mechanical vibration – Evaluation of machine vibration by measurements on non-rotating parts – Part 1: General guidelines [1]. A sample machine supported on sliding bearings is presented in Figure 1.



Fig. 1. Motor of a mill in a cement plant, that is three-phase induction slip-ring motor DOLMEL SYUe-148r/01 type, with power of $P_N = 1000$ kW, $n = 738$ rpm.

2 Measurements of vibrations of electric machines

A well designed and performed machine is characterised by a low level of vibrations. In the course of operation, a machine wears out, its foundations subside, machine elements become deformed and slight changes of the dynamic qualities of a machine appear with time. What occurs is excessive non-coaxiality of shafts linked to clutches, wear of parts, increased imbalance of vibrating elements and increased play. All these factors are reflected in the rise of energy of machine vibrations [2]. Resonance vibrations and significantly increased dynamic loads are caused as a result. A vibrating signal carries a lot of information connected with the technical condition of electric motors and is a basis for these signals to be applied in on-line monitoring systems as indicators of machine condition trend. The spectral analysis of these signals allows to identify the type of damage [2], [3].

2.1 Measurement methods

The following values can function as vibration measurement parameters:

- vibration relocation, measured in micrometres;
- vibration speed, measured in millimetres per second;
- vibration acceleration, measured in metres per second squared.

It should be taken care of for vibrations to be measured in a wide band encompassing the scope of a machine frequency spectrum. A range from 10 Hz to 1000 Hz is

frequently assumed. However, different frequency ranges and different measurement values may correspond to different types of machines. Vibration measurements are carried out in various measurement points, in one, two or three directions. The maximum broadband value of spot vibration intensity is called the intensity of machine vibrations. Measurement points of absolute rotations of electric engines are located on the casings of bearings and their supports, non-rotating structural elements carrying dynamic loads and characteristic of the vibrating condition of a machine. Typical spots for the location of vibration transducers are shown in Figure 2.

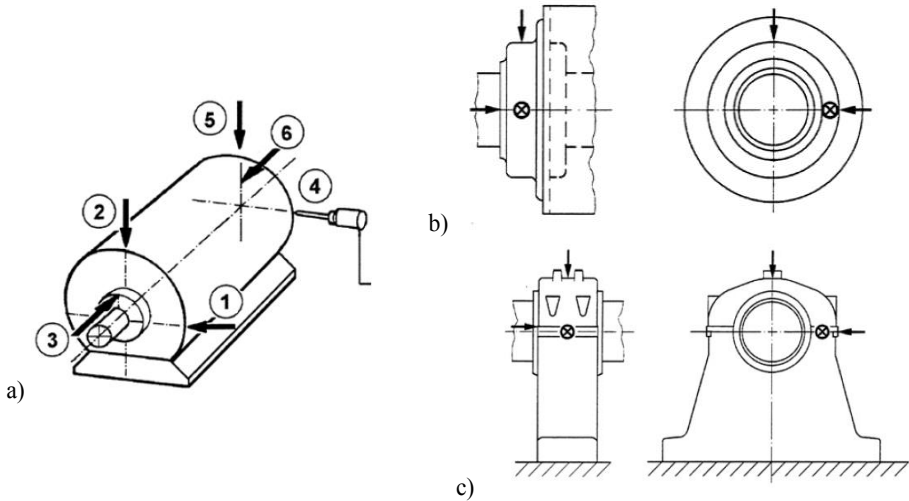


Fig. 2. Recommended location of transducers for measurement of absolute vibrations: a) for small electric machines, b) on bearings in casings, c) on free-standing bearings.

In the process of making a decision concerning the type and size of a transducer of absolute vibrations, as well as the location and manner of mounting it, the following rules should be observed:

- a) the weight of the installed transducer cannot effect a change of vibration characteristics of the entire structure, that is it cannot exceed 1/10 of the vibrating element weight;
- b) the dynamics and the applied frequency range of a transducer cannot be limited by too low stiffness of its mounting as this may lead to deformation of measurement results of vibrations with frequencies higher than 500 Hz;
- c) prior to assembling transducers, it is recommended to carry out initial measurements in different parts of a machine in order to determine the most appropriate location of measurement spots;
- d) preparing and marking measurement spots in advance is likely to increase measurement precision and repeatability of vibrations caused by transducers which are not permanently fixed.

2.2 System for measuring absolute rotations

Five low-profile accelerometers of SKF CMSS-2200 type, with a constant of 100 mV/g, were mounted in the tested engine, Figure 3:

- accelerometer 1 – bearing stand, drive end, vertical direction;
- accelerometer 2 – bearing stand, drive end, horizontal direction;
- accelerometer 3 – bearing stand, non-drive end, horizontal direction;
- accelerometer 4 – bearing stand, non-drive end, vertical direction;

Acceleration sensors were mounted on the stands of sliding bearings with the use of a screw connection, which assures the broadest band for transmitting vibrations. Among other methods of mounting transducers, there is manual application, attaching them with the use of wax or glue and a magnetic grip. The sensors used for measuring absolute vibrations were connected to the SKF analyser with the use of a special screen cable with a double shielding layer CMSS 9200.

a)



b)



Fig. 3. Measurement accelerometers a) mounted on a bearing stand of an engine, on the drive end, b) sample acceleration meter used in the measurement system – CMSS 2200 [5].

Absolute vibrations are measured in three reciprocally perpendicular directions: vertical, horizontal and axial. Such a measurement makes it possible to test values such as acceleration (a), speed (v) and relocation (s). Selecting a proper measurement value depends on, first of all, the qualities of the tested electric motor. Measuring acceleration is the best measure to reflect vibrations of a machine with low frequencies of about 10 Hz and slow-rotation machines whose rotary speed does not exceed 600 rpm. This is because low frequencies of vibrations are characterized by considerable relocations. Vibration speed is the best way to represent the condition of a machine with frequencies within the range from 30 Hz to 1500 Hz, whereas drive acceleration, is best for high frequency machines [4].

3 On-line monitoring of absolute vibrations

The existing critical drive with power of $P_N = 1000$ kW and rotary speed of $n = 738$ rpm, applied in one of Polish cement plants was used for diagnostic testing. @plitude Observer software by SKF company was used for visualisation of measurement results. The highest registered vibrations in the assumed period of time for the spectral diagram of a sliding bearing of the tested electric motor were shown in Fig. 4.

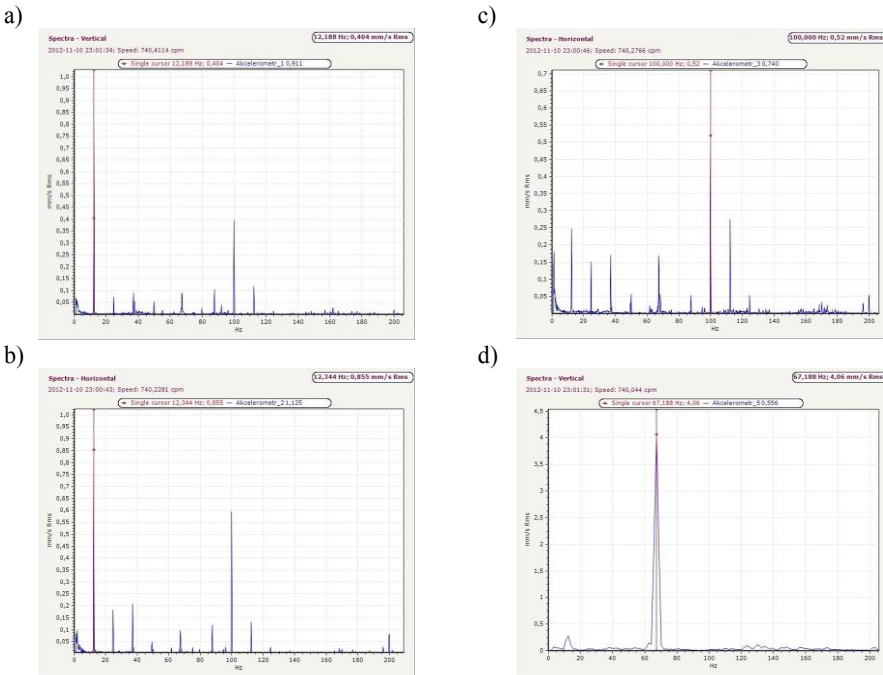


Fig. 4. Registering the highest values of absolute vibrations for the spectral diagram of a sliding bearing a) accelerometer 1, drive end, vertical direction; b) accelerometer 2, drive end, horizontal direction; c) accelerometer 3, non-drive end, horizontal direction; d) accelerometer 4, non-drive end, vertical direction;

Frequency analysis of absolute vibrations consists in studying the characteristic bands appearing in the spectral diagram and corresponding to the damage that occurs. From the above diagrams we can read that on 10th November 2012, at 11:00 pm., the following vibration speeds on drive end were registered: for the sensor measuring in vertical direction 0.404mm/s at the frequency of 12.188 Hz, and for the sensor measuring in horizontal direction 0.855 mm/s at the frequency of 12.344 Hz. In the same measurement period, the following vibration speeds on non-drive end were registered: for the sensor measuring in vertical direction 0.52mm/s at the frequency of 100 Hz, and for the sensor measuring in horizontal direction 4.06mm/s at the frequency of 67.188 Hz. Taking into consideration the fact that the tested electric engine belongs to class III – large driving units and other large machines with vibrating

weights, mounted on fixed and heavy foundations with low vibration measurability and the classification of vibration intensity in accordance with the ranges shown in Table 1, we can perform current valuation of the technical condition of a machine.

Table 1. Typical limit values of effective vibration speeds, corresponding to individual zones for class III of machine classification, in accordance with PN-ISO-10816-1:1998 standard.

Zone name	Bottom values of the zone RMS [mm/s]	Upper values of the zone RMS [mm/s]
<i>A Zone</i> – machines directly after technical acceptance	0.28	2.80
<i>B Zone</i> – machines approved for long-term operations without any operational restrictions	2.80	7.10
<i>C Zone</i> – machine is not suitable for long-term operation, taking remedial measures in necessary	7.10	18.0
<i>D Zone</i> – machine may be damaged in the course of further operation	18.0	45.0

By carrying out the frequency and spectral analyses of the obtained results, we can conclude that the condition of sliding bearings on the drive end is within the A zone, whereas those on the non-drive end are in the B zone. The drive is fully operational in case of both bearings.

3 Summary

Online monitoring of absolute vibrations is a good, frequently applied and relatively easy manner of analysing technical condition of the supervised electric motors used in industry. One should remember that vibration measurement is limited due to the factors influencing signal passing. Among the most important ones there are: internal machine dampening, phase delay of a signal, structural resonances and, most of all, the influence of oil film in the bearing, which is characterised by high non-linearity.

References

1. PN-ISO 10816-1 Mechanical vibration – Evaluation of machine vibration by measurements on non-rotating parts – Part 1: General guidelines.
2. Cempel Cz.: Vibroacoustic Diagnostics of Machines; PWN, Warszawa, Poland 1989,
3. Szymaniec S.: Testing, Operation and Diagnostics of Machine Sets with Squirrel-cage Induction Motors; Printing House of Opole University of Technology, Opole, Poland 2013, Studies and Monographs, paper 333.
4. Dwojak J., Rzepiela M.: Vibration Diagnostics of Machine and Appliances Condition, 2nd edition; Gamma Office, Warszawa, Poland 2005.
5. SKF Vibration Sensors Catalog PUB CM/P1 11604/8 EN, June 2013.

Analysis of Iron and Copper Losses in Permanent Magnet Synchronous Motor

Piotr Mynarek

Opole University of Technology,
Faculty of Electrical Engineering, Automatic Control and Computer Science,
Institute of Electromechanical Systems and Industrial Electronics
ul. Prószkowska 76, 45-758 Opole, Poland
piotr_mynarek@o2.pl

Abstract. This paper presents the calculation of copper and iron losses in the permanent magnet synchronous motor (PMSM). Calculations are performed using a three-dimensional field-circuit finite-element numerical model. In order to validate correctness of the calculation results, measurements on the physical model of the machine are carried out.

1 Introduction

In recent years synchronous motors with permanent magnets (PMSM) are gaining more popularity in drives dedicated for continuous operation such as pump and fan [1, 3]. These machines, by using neodymium magnets, are characterized by high efficiency, high power density, reliability and a wide range of motor speed. Additionally PMSM motors are characterized by a high overload torque, combined with a relatively low moment of inertia of the rotor provides good dynamic properties [2]. One of the most important aspects in the design of PMS motors is to determine the losses occurring in the motor. This analysis allows to determine the efficiency and assess of heating machines. Otherwise increase the power output of an electrical device is dictated mainly by its thermal behavior. In addition, in the case of permanent magnet motors thermal analysis is extremely important because of the risk of a partial demagnetization of permanent magnets as a result of too much of their warm-up. Therefore, works connected with determining of power losses occurring in motors with permanent magnets in various phases of their work become appropriate. This article attempts to analyze iron losses in the PMSM.

2 Electromagnetic Model of PMSM

The work adopted a synchronous motors with permanent magnets Sh80-6B SMKwsg90M8 as the object of research. The motor parameters:

- rated output $P_n=2\text{kW}$,
- rated current $I_n=3,7\text{A}$,

- rated voltage $U_n=400\text{V}$,
- rated speed $n_n=2200\text{obr/min}$,
- efficiency $\eta=91\%$.

In order to determine the losses in the test PMSM was built field-circuit model in Flux3D environment, based on the finite element method (FEM). The complexity of the field-circuit model require the use of sophisticated algorithms and computers with high calculating powers. Despite continuous improvement in both areas, the existing software does not provide the user with the freedom of action due to software limits and high hardware requirements. Therefore, it is necessary to make the necessary assumptions, and limiting the complexity of the task, and at the same time ensure a correct solution. Therefore, a natural approach to the construction field models of electrical machines is the use of boundary conditions allow to minimize the numerical cost. In designed field model of the analyzed motor, were used the periodic and the symmetry boundary conditions, thereby reducing the calculations to 1/8 volume of the entire machine. During the analysis the following simplifications were made: hysteresis was ignored as well as eddy currents, constant density of coil current was assumed.

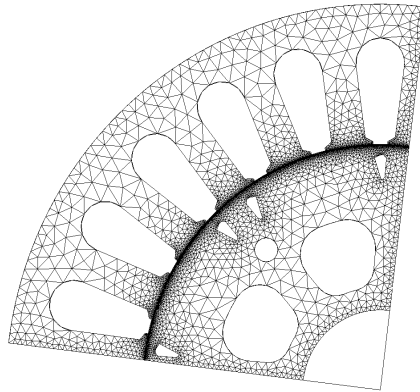


Fig. 1. Discretization mesh of the PMS motor

3 Results of analysis

Based on field-circuit model, a number of computer simulations were performed. In order to verify the correctness of the numerical model in the first stage of the study, the calculations the phase current waveforms in steady state for different operating conditions were performed. In Figure 2a shows a computer simulation of currents waveforms in the phase for machines loaded torque $T_l = 0.46 \text{ N}\cdot\text{m}$ at nominal speed and voltages $U_f = 265\text{V}$. Figure 2b shows current waveforms with an motor load torque $T_l = 8.7 \text{ N}\cdot\text{m}$.

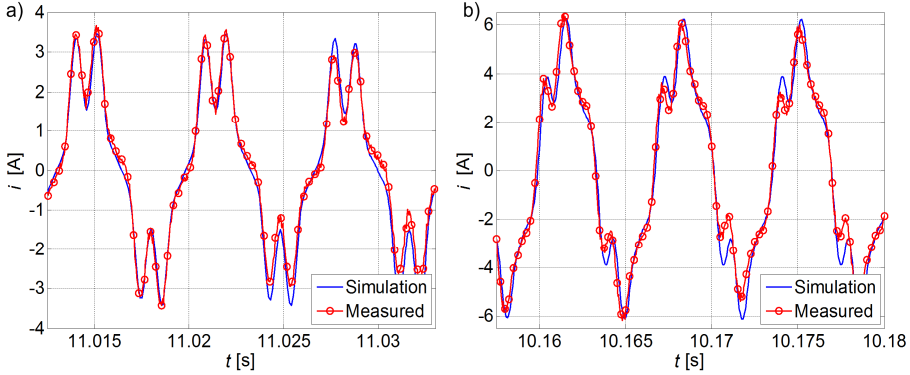


Fig. 2. The phase current waveforms in steady state

Power losses in the copper are calculated according to the formula:

$$\Delta P_{Cu} = RI^2 \tag{1}$$

where R is the phase resistance, I is phase current.

Since the object of study is the synchronous motor, iron losses are mainly formed in the stator. In the case of the rotor ΔP_{Fe} losses are due to the presence of harmonic. But these losses globally are negligibly small and therefore the study the only losses in the stator. The calculation of iron loss is based on the modified Steinmetz equation that accounts for hysteresis, classical Joule eddy-current, and excess losses[4].

$$\Delta P_{Fe} = \iiint_v (k_h B_m^2 f k_f + \frac{1}{T} \int_0^T \left[\sigma \frac{d^2}{12} \left(\frac{dB}{dt}(t) \right)^2 + k_e \left(\frac{dB}{dt}(t) \right)^{\frac{3}{2}} \right] k_f dt) dv \tag{2}$$

where k_h is the hysteresis coefficient, B_m maximum flux density excursion over an electrical cycle, f is the frequency, k_f is the packing factor of the laminated core pack, σ is the electric conductivity, d is the thickness of a single sheet of lamination, and k_e is the excess loss coefficient, v is the volume of the region. The iron losses is calculated within each element of the meshed core region using (1) and then averaged over the entire stator core volume. Table 1 lists the material coefficients used for the iron losses predictions.

Table 1. Material coefficients for M470-50A lamination

Coefficients	Symbol	Value	Unit
Hysteresis coefficient	k_h	143	[Ws/T ² /m ³]
Excess loss coefficient	k_e	2,6	[W/(Ts ⁻¹) ^{3/2} /m ³]
Packing factor	k_f	0,96	-
Density	σ	3,03 · 10 ⁶	[(Ωm) ⁻¹]
Lamination sheet thickness	d	0,5 · 10 ⁻³	[m]

Copper and iron losses were determined for the two specific areas of the stator: stator yoke and the stator teeth (Table 2).

Table 2. Calculated loss in PMSM

Load on the motor	ΔP_{Cu} [W]	ΔP_{Fe} [W]	
		Stator teeth	Stator yoke
$T_f=0,46\text{N}\cdot\text{m}$	17,11	36,31	17,14
$T_f=8,7\text{N}\cdot\text{m}$	64,89	40,6	21,82

4 Summary

The paper presents the results of the field-circuit analysis to determine the copper and iron losses in the permanent magnet synchronous motor. In order to validate correctness of the calculation results, measurements on the physical model of the machine are carried out. With the use of the constructed model one may fast and at the same time with sufficient precision determine iron losses. This approach allows a better determination of heat sources present in the motor needed to thermal analysis of the machine. Thermal analysis of the test motor is planning in the further part of the work.

References

1. Barański M., Szelağ W., Jędryczka C., Mikołajewicz J.: Analysis of line start permanent magnet synchronous motor with u-shaped magnets rotor. *Prace Naukowe Instytutu Napędów i Pomiarów Elektrycznych Politechniki Wrocławskiej* Nr 66, Wrocław 2012, s.78-85.
2. Król E.: Comparison of energy efficiency of permanent magnet synchronous motors and induction motors. *Zeszyty Problemowe – Maszyny Elektryczne* Nr 78/2007, wyd. BOBRME Komel s.75-78.
3. Wrobel R., Mellor P. H.: Thermal Design of a High Energy Density Wound Components, *IEEE Transactions on Industrial Electronics*, Vol. 58, No. 9, 2011, pp. 4096-4104.
4. Zalas P., Antal L.: Low power permanent magnet synchronous motor designed to synchronous starting. *Zeszyty Problemowe – Maszyny Elektryczne* Nr 3/2012, wyd. BOBRME Komel, s. 125-130.



Piotr Mynarek is a recipient of a Ph.D. scholarship under a project funded by the European Social Fund

Evaluation of algorithms of SIFT, BRIEF, FREAK descriptors, and their effectiveness when comparing images undergoing geometric translations

Marcin Niedźwiecki

Opole University of Technology, Institute of Control and Computer Engineering,
Faculty of Electrical Engineering, Automatic Control and Informatics,
Opole, Poland

Abstract. Currently, the key problem in Computer Vision applications is efficient and effective approach to generating keypoints of images, and creation of descriptors for later comparison and matching. In this article I will try to bring the most popular and effective SIFT, BRIEF methods, and confront them with new methods appearing on the horizon, such as FREAK. And to consider in the end which of the algorithms will be the best during the times of an increasing emphasis on the deployment of these solutions on such mobile devices as smartphones and tablets and other embedded devices that impose restrictions in terms of memory capacity and computing power and focus on performance in real-time applications at the same time.

Keywords: SIFT, BRIEF, FREAK, geometric translations

1 Preamble

Object recognition and matching, 3D reconstruction, and motion tracking – all rely on the presence of stable, representative features in the image.

The ideal keypoint detector finds salient image regions such that they are repeatably detected despite possible image translations.

The main objective of teams dealing with this issue is to create an algorithm running as efficiently as *Scale Invariant Feature Transform* (SIFT) proposed in 1999 by David G. Lowe.

The search is on algorithms with lower computation requirements and a maintained invariance to the most popular image translations (e.g. scaling, rotation, geometric distortions ...).

The works to date resulted in the creation of methods that meet the above requirements at least in principle i.a. such solutions as the following:

- *Speeded Up Robust Features* (SURF) [2]
- *Binary Robust Independent Elementary Feature* (BRIEF) [3]
- *Binary Robust Invariant Scalable Keypoints* (BRISK)[4]
- *Fast Retina Keypoints* (FREAK)[1]

2 Overview of solutions

2.1 Scale Invariant Feature Transform (SIFT)

Introduction The approach proposed in 1999 by David G. Lowe, a Canadian professor working at the University of British Columbia, called the *Scale Invariant Feature Transform* (SIFT) [1] has become a generally accepted model, due to its effectiveness. The descriptor provides invariance to most of the most popular image translations. Unfortunately, SIFT method requires a lot of computing power.

This algorithm is used for the following:

- object recognition, robotic mapping and navigation, image stitching, 3D modeling, gesture recognition, video tracking, match moving.

The algorithm was patented in the United States and is owned by the University of British Columbia.

The mechanisms of the algorithm First, the keypoints of objects are extracted from a set of reference images, and stored in a database.

An object is recognised in a new image by individually comparing each feature from the new image to this database. Then, a set of keypoints is sought, whose mutual Euclidean distances are as close as possible.

The obtained set of keypoints is filtered for compatibility according to the location, scale, and orientation in the new image in order to filter out good matches.

The determination of consistent clusters is performed rapidly by using an efficient method based on hash table. The table implements generalised Hough transform methods – methods of extracting regular shapes.

2.2 Binary Robust Independent Elementary Features (BRIEF)

A milestone, designed to create a solution tailored to devices with limited resources, was to develop a method to create an image descriptor called *Binary Robust Independent Elementary Features* (BRIEF).

BRIEF algorithm in combination with the method for detecting the keypoints called *Features from Accelerated Segment Test* (FAST), allowed to apply it in real-time applications.

The authors have proposed the use of binary strings as an efficient feature description. Thanks to a range of treatments, BRIEF is a very efficient algorithm for creating, comparing and matching features.

Authors of the paper [2] tested their algorithm with standard benchmarks, and compiled the results with other rapid methods such as SURF and U-SURF. Analysis of these data shows that the fruit of their work achieves similar or sometimes higher effectiveness in a shorter time and with lower computational requirements.

Unfortunately BRIEF is characterised by a low tolerance to image distortions and noise, especially to rotation and change of scale.

Table 1. Problems and solving methods, and the benefits of using SIFT algorithm.

Problem	Technique	Advantages
Key localization / scale / rotation	DoG(Difference of Gaussians) / scale-space pyramid / orientation assignment	accuracy, stability, scale & rotational invariance
Geometric distortion	blurring / resampling of local image orientation planes	Affine invariance
Indexing, matching	Nearest neighbours / Best Bin First search	Efficiency and speed
Cluster identification	Hough Transform voting	reliable models pose
Model verification / outlier detection	Linear least squares	better error tolerance with Fewer Matches
Hypothesis accep- tance	Bayesian Probability analysis	reliability

2.3 Fast Retina Keypoints (FREAK)

The motivation for the authors of this algorithm was to achieve the best possible results based on solutions which are suggested by the human body while maintaining the computational complexity as low as possible. They were inspired by the human retina.

Topology of the human retina proved to play so important role for the authors that they decided to transpose its structure, and design image descriptor on its basis.

Based on the progress of medicine in understanding of the methods of how the image is transmitted to the brain, i.e. the assertion that the human retina extracts details from images using Difference of Gaussians, and encodes such differences with action potentials, the authors decided to use these methods in their own algorithm.

3 Tests and results

3.1 Introduction

The comparative test for descriptors operation was carried out for three methods recognised by me as the most interesting, i.e. SURF, BRISK and FREAK.

The sequence of testing operations is as follows:

- Conversion of input image to grayscale

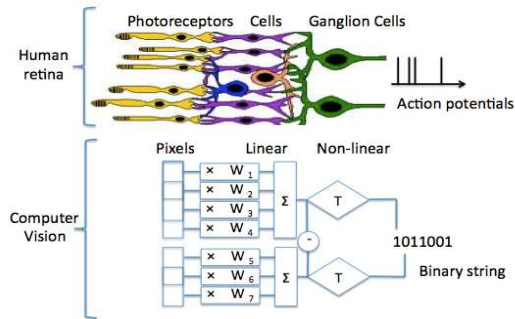


Fig. 1. From human retina to computer vision: the biological pathways leading to action potentials are emulated by simple binary tests over pixel regions. (Upper part of the image is a courtesy of the book *Avian Visual Cognition* by R. Cook.)

- Detection of keypoints and extraction of descriptors from input grayscale image
- Generation of all transformed images using passed transformation algorithm
- Then, for each of the transformed images:
 - Detection of keypoints and extraction of descriptors
 - Match of train descriptors and query
 - Computation of statistics for the testing procedure

Testing conditions for computing platform The whole process of testing was conducted using OpenMP (called Open Multi-Processing) – an application programming interface (API) that supports multi-platform shared memory multiprocessing programming. Tests were conducted on a MacBook Pro laptop equipped in the following components:

- The processor in the equipment used is a Quad Core i7, which loaded all cores for 100% during the tests, 16GB RAM, 256Mb SSD

3.2 Results

Results are shown in Figures 2 to 4.

4 Conclusions

In a surprising way, the results have shown how low is the level of matching ratio for SURF in the case of testing images that have different rotation with respect to each other.

Based on the above results we can say that BRISK algorithm has proved to be very positive. It is characterised by a fantastic effectiveness at two times faster operation. Taking into account the results of the tests, I think that from

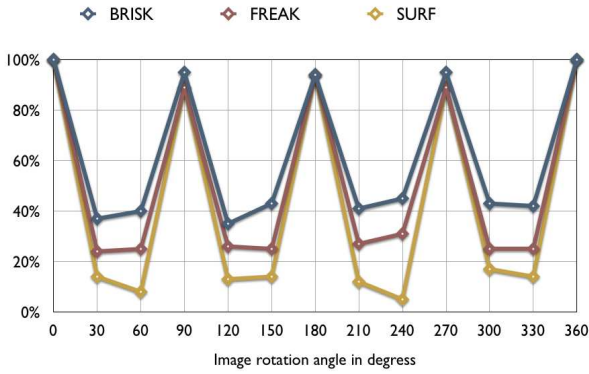


Fig. 2. The effectiveness of the algorithms depending on the rotation

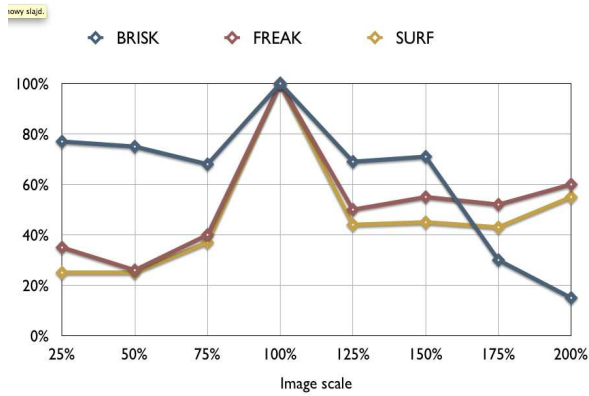


Fig. 3. The effectiveness of the algorithms depending on the scale

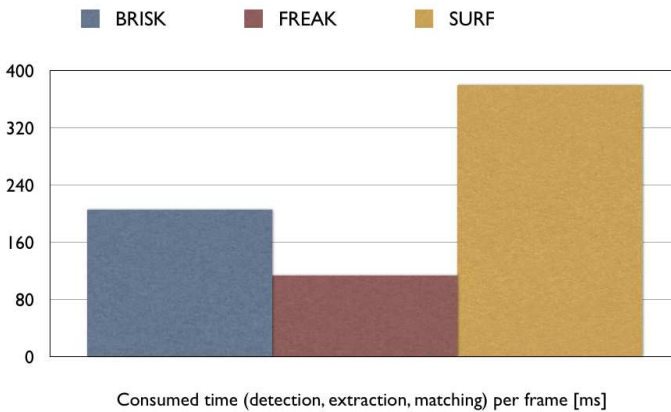


Fig. 4. The time required for analysis and comparison of a single frame

the group of presented algorithms BRISK (and some cases FREAK) is a natural candidate to be used in applications focused on efficient and effective analysis and image matching in real-time applications.

References

1. FREAK: Fast Retina Keypoint – Alexandre Alahi, Raphael Ortiz, Pierre Vandergheynst Ecole Polytechnique Federal de Lausann (EPFL), Switzerland
2. SURF: Speeded Up Robust Features – Herbert Bay, Tinne Tuytelaars, Luc Van Gool, ETH Zurich
3. BRIEF: Computing a Local Binary Descriptor Very Fast – M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C.Strecha, and P.Fua, IEEE Transactions on Pattern Analysis and Machine Intelligence
4. BRISK: Binary Robust Invariant Scalable Keypoints – Stefan Leutenegger, Margarita Chli and Roland Y. Siegwart Autonomous System Lab, ETH Zurich

Implementation of the Plume Rise Procedure in the Massively Parallel Architecture

Tomasz Przybyła

Opole University of Technology,
Faculty of Electrical Engineering, Automatic Control and Computer Science,
Institute of Automatic Control and Computer Science,
ul. Sosnkowskiego 31, 45-272 Opole, Poland
t.przybyla@doktorant.po.edu.pl

Abstract. This paper presents the performance improvement achievable in Gaussian Puff Model through parallelization of the plume rise procedure. Calculations carried out in the massively parallel architecture allowed to accelerate the calculations associated with the appointment of transitional momentum plume rise.

1 Introduction

In air quality management system the assessment of air pollution is performed on the basis of measurement systems and mathematical modeling. There are many models for simulate the spread of pollutants in ambient air. Starting from simple box models to complex models used for large-scale and different meteorological conditions [1, 2]. One of the universal tools recommended by the Ministry of Environment in Poland is CALPUFF (CALifornia PUFF Model) [3].

2 CALPUFF model

CALPUFF is a universal model which can be used to model the spread of various kinds of gaseous and particulate pollutants in ambient air. In this model the plume of pollutions emitted from the source is approximated by a series of the puffs which transport and shape is determined by the weather conditions [2, 3, 4]. CALPUFF model allows to determine the concentrations of pollutants coming from various emission sources: point, line, surface and volume [3].

In the CALPUFF calculation are performed in the three grids: meteorological, computational and receptors. The concentration of pollutants at the reception point is determined by adding the concentrations of all puffs. Whereas the number of generated puffs depends largely on weather conditions. In the CALPUFF model calculations are performed in three main loops: hours loop, puffs loop and receptors loop [3].

3 Plume rise procedure

One of the procedures of the CALPUFF model is PRM (plume rise momentum) procedure. The PRM is called by the GRISE procedure, which is responsible for the calculation of height of gradual plume rise at a given point. The plume rise algorithm takes into account phenomena such as plume buoyancy and momentum, vertical wind shear, stable atmospheric stratification, area sources plume rise and line source plume rise [3]. The plume rise is estimated by the formula (1).

$$z_n = \left[\left(\frac{3F_m x}{\beta_j^2 u_s^2} \right) + \left(\frac{3F x^2}{2\beta_i^2 u_s^3} \right) \right] \quad (1)$$

$$\beta_j = \frac{1}{3} + \frac{u_s}{w} \quad (2)$$

where: F_m – the momentum flux,
 F – the buoyance flux,
 u_s – the stack height wind speed,
 x – the downwind distance,
 β_i – the neutral entrainment parameter,
 β_j – the jet entrainment coefficient,
 w – the stack gas exit speed.

The distance in which the plume reaches a maximum elevation (3)

$$x_f = f(x) = \begin{cases} 3.5x^*, & F > 0 \\ \frac{4D(w + 3u_s)^2}{u_s w}, & F = 0 \end{cases} \quad (3)$$

$$x^* = f(x) = \begin{cases} 14F^{\frac{5}{8}}, & F \leq 55 \\ 34F^{\frac{2}{5}}, & F > 55 \end{cases} \quad (4)$$

where: D – the stack diameter.

This routine is used for calculate transitional momentum plume rise for neutral and stable condition.

4 Parallelization

To speed up the calculations used the massively parallel architecture of the GPU (Graphics Processing Unit) and OpenCL standard. Implementation of parallel computing in OpenCL starts from defining the problem in the N-dimensional space. Each

independent component in the index space is called a work-item. Each work-item performs the same program kernel, but for different data [5, 6, 7].

The calculation was carried out on the following platform: software: Visual Studio 2010, NVIDIA CUDA SDK 5.0, Intel OpenCL SDK 1.5; hardware is as shown in Table 1.

Table 1. Comparison of computational resources

	Host	GPGPU
Device	Intel Core i7-2630QM	NVIDIA GT 540M
Platform	Intel(R) OpenCL 1.1	OpenCL 1.2 CUDA 5
Compute Units	8	2
Processing Elements	8	96
Frequency	2000 MHz	1600 MHz

Each processing element (PE) determines the value of PRM function for a given point of a grid.

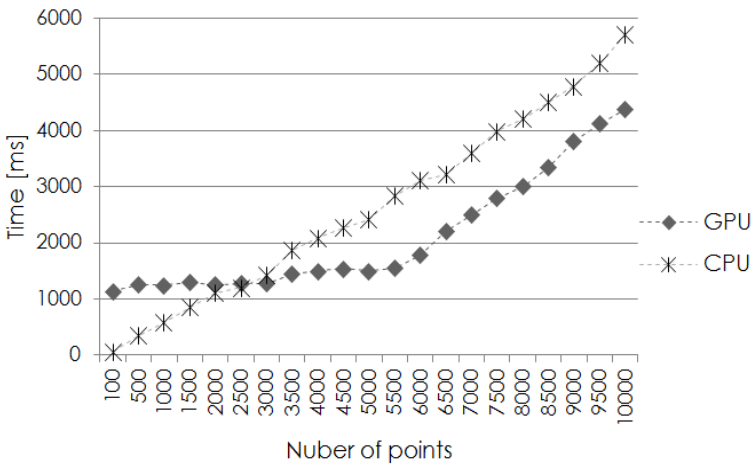


Fig. 1. Results of the computation time between the CPU and GPU

Performance results show that OpenCL has good potential to be speed up one of the procedure of CALPUFF model. For a small number of points experimental results show that OpenCL implementation on GPU achieves a longer computation time. In this case, the initialization time of the platform and transfer the data to a computing device significantly affects on the total computation time. The OpenCL implementation gives better performance when the number of independent tasks is sufficiently large and do not need frequent references to main memory.

5 Summary

The results show that the implementation of the *PRM* procedure in the massively parallel architecture enabled to accelerate the calculations associated with the appointment of transitional momentum plume rise. However, as shown by the results, reducing computation time is remote. This is due to of the size of task and the need to copy data between the CPU and GPU memory.

References

1. Łobocki L.: Methodological Guidance of the Mathematical Modeling in the Air Quality Management, Ministry of Environment, General Inspectorate for Environmental Protection, Warsaw, 2003 (in Polish).
2. Markiewicz M. T.: Fundamentals of Modeling the Spread of Pollutants in the Atmosphere, Warsaw University of Technology, Warsaw, 2004 (in Polish).
3. Scire J.S., Strimaitis D.G., Yamartiono R.J.: A User's Guide for the CALPUFF Dispersion Model (Version 5), Earth Tech, Inc. Concord, 2000.
4. Alexandrov V.N., Owczarz W., Thomson P. G., Zlatev Z.: Parallel runs of a large air pollution model on a grid of Sun computers, *Mathematics and Computers in Simulation* 65, pp. 557-577, 2004.
5. Gaster B., Howes L., Kaeli D., Mistry P., Schaa D.: *Heterogeneous Computing with OpenCL*. Elsevier Science & Technology, 2011.
6. Munshi A.: *The OpenCL Specification version 1.2*. Khronos OpenCL Working Group, 2011.
7. Przybyła T.: Calculations in the massively parallel architecture in heterogeneous systems, XIII International PhD Workshop, OWD, Wisła 2011, s. 30-34.



Tomasz Przybyła is a recipient of a Ph.D. scholarship under a project funded by the European Social Fund

Predictive algorithms for three-wheeled mobile robot control

Tomasz Pyka

Opole University of Technology,
Institute of Electromechanical Systems and Industrial Electronics,
Proszkowska Street 76, 45-768 Opole, Poland
t.pyka@doktorant.po.edu.pl

Abstract. In this paper the author compares the predictive method for determining the duty cycle rate of modulation (PWM) with the predictor-corrector method. The use of these methods could reduce the error on the trajectory of the mobile robot. The simulations were carried out for two different process control steps and it was assumed that the pulse width modulation (PWM) frequency was equal to 5 kHz..

1 Introduction

One of the most frequently discussed issues in the literature is precise control of autonomous mobile robots. Currently, the algorithms used are robust or adaptive [1, 2], which significantly reduces the errors. It should also be noted that when more sophisticated control algorithms are used, they need more computing power. Another problem is a relatively low frequency carrier wave of PWM, which in the DC motor control in mobile robots, currently rarely exceeds 10kHz. In practice, the appropriate choice of duty cycle rate of modulation is a prerequisite for correct control.

Therefore, when design of the robot does not allow for complex control algorithms, it still uses predictive methods which can be written in a simple form and do not require complex computing power. In that case, two algorithms were compared: a predictor Adams-Bashforth 1st and 2nd order algorithm and Adams-Moulton predictor-corrector algorithm of the 2nd and 3rd order.

The aim of this paper is to answer which of the algorithms is more suitable for practical applications with computing power absorption being reduced at the same time. Both methods were compared designating the duty cycle to PWM rate without prediction.

The results were obtained by computer simulation of a three wheeled mobile robot Pioneer 2DX [3].

2 Mobile robot model

Model of mobile robot and equations are taken from [3] and written in the following form:

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{h}(\mathbf{q}, \dot{\mathbf{q}}) = \mathbf{T}, \tag{2}$$

where:

$$\mathbf{M}(\mathbf{q}) = \begin{bmatrix} 2m_1l_1^2 + m_5l_2^2 + I_{z5} + 2I_{y1}h_1^2 \\ (2m_1 + m_5)r_1^2 + 2I_{z1} \end{bmatrix}, \quad \mathbf{h}(\mathbf{q}, \dot{\mathbf{q}}) = \begin{bmatrix} -m_5l_2r_1\dot{\psi}\dot{\beta} \\ m_5l_2r_1\dot{\beta}^2 \end{bmatrix},$$

$$\mathbf{T} = \begin{bmatrix} (M_1 - M_2 - N_1f_1 + N_2f_2)h_1 \\ M_1 + M_2 - N_1f_1 - N_2f_2 \end{bmatrix}, \quad \mathbf{q}^T = [\beta \quad \psi],$$

where \mathbf{M} and \mathbf{C} describing robot dynamics, \mathbf{T} consist forces and generalized forces, m_1 – mass of the drive wheel, m_5 – mass of construction, r_1 – radii of the drive wheel, l_1 – distance between the center of the frame mass and the characteristic point, l_2 – distance from the characteristic point to the driving wheel, M_1, M_2 – torques, N_1, N_2 – drive wheel pressure of the ground, f_1, f_2 – friction coefficient of the ground, $h_1 = l_1/r_1$, I_{z1}, I_{z5}, I_{y1} – moments of inertia, ψ – rotation angle of the substitute drive wheel, β – rotation angle of the frame.

The simulation was carried out using the real DC motor Dunkermotoren GR63 parameters with gear PLG52 1:20.25, where: $U_n=24.0V$, $I_n=2.7A$, $R_t=1.2 \Omega$, $L_t=0.0029$ H, where: U_n – rated voltage, I_n – nominal current, R_t – armature resistance, L_t – armature inductance.

In the simulation the author included a simplified model of the inverter voltage which was presented in [4].

2 Predictor and predictor-corrector algorithms

Below the author compares the predictor Bashforth algorithm with the Adams Moulton predictor-corrector algorithm for two different time steps of the control process. It is assumed that the PWM frequency is 5 kHz and the step process control step of duty cycle rate is $t \in \{0.002s; 0.02s\}$. For the 1st and 2nd order, the author analyzes Bashforth method, and for the 2nd and 3rd order – Adams-Moulton method.

Therefore, the algorithms are written below:

1st order Adams-Bashforth [5]:

$$x_{n+1} = x_n + hf(x_n, t_n), \tag{3}$$

2nd order Adams-Bashforth [5]:

$$x_{n+1} = x_n + h \left\{ \frac{3}{2} f(x_n, t_n) - \frac{1}{2} f(x_{n-1}, t_{n-1}) \right\}, \tag{4}$$

2nd order Adams-Moulton [5]:

$$x_{n+1} = x_n + h \left\{ \frac{1}{2} f(x_{n+1}, t_{n+1}) + \frac{1}{2} f(x_n, t_n) \right\}, \quad (5)$$

3rd order Adams-Moulton [5]:

$$x_{n+1} = x_n + h \left\{ \frac{5}{12} f(x_{n+1}, t_{n+1}) + \frac{8}{12} f(x_n, t_n) - \frac{1}{12} f(x_{n-1}, t_{n-1}) \right\}. \quad (6)$$

3 Simulation results

Simulation results are shown in Table 1:

Table 1. Simulation results

Algorithm	Process control step	Root mean square error [m]
without prediction	t=0.002s	$1.84 \cdot 10^{-1}$
without prediction	t=0.02s	$1.85 \cdot 10^{-1}$
Adams-Bashforth 1 st order	t=0.002s	$1.69 \cdot 10^{-1}$
Adams Bashforth 1 st order	t=0.02s	$1.59 \cdot 10^{-1}$
Adams Bashforth 2 nd order	t=0.002s	$1.12 \cdot 10^{-1}$
Adams Bashforth 2 nd order	t=0.02s	$1.27 \cdot 10^{-1}$
Adams Moulton 2 nd order	t=0.002s	$1.69 \cdot 10^{-1}$
Adams Moulton 2 nd order	t=0.02s	$1.59 \cdot 10^{-1}$
Adams Moulton 3 rd order	t=0.002s	$1.39 \cdot 10^{-1}$
Adams Moulton 3 rd order	t=0.02s	$1.31 \cdot 10^{-1}$

Based on the results shown in Table 1, it can be concluded that the predictor method of Adams-Bashforth for the 2nd order is sufficient for use in case when the computing power of microcontrollers is low. Errors obtained by using the Adams-Moulton predictor-corrector algorithm are to a significant extent comparable with the results obtained after using the Adams-Bashforth algorithm. The results for the 2nd order Adams-Moulton algorithm and the 1st order Adams Bashforth algorithm are equal. The method without the prediction has the same errors for both process control step.

Figures 1 and 2 show the simulation results:

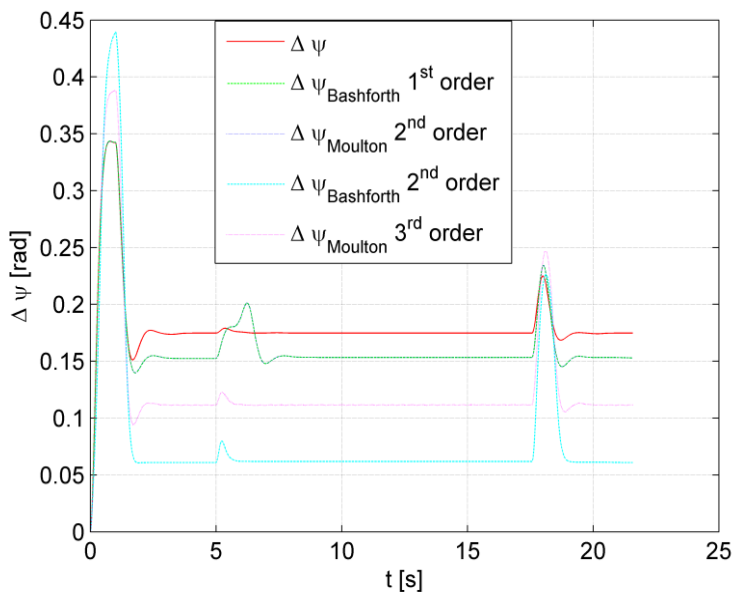


Fig. 1. Rotation angle ψ errors for timestep process $t=0.002s$

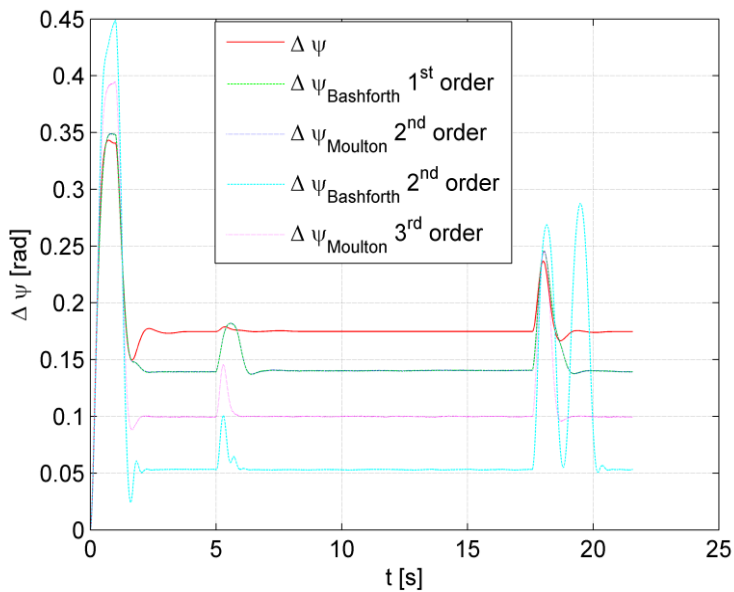


Fig. 2. Rotation angle ψ errors for timestep process $t=0.02s$

When comparing figures 1 and 2, we can see that for the process control step $t = 0.02s$, in case of the Adams-Bashforth method of the 2nd the area of stability has been violated, and thus at the end of the robot motion trajectory oscillations occurred (17-20s).

4 Summary

Based on the simulations presented above it can be concluded that for industrial tasks, when computing power is limited, the Adams-Bashforth algorithm of the 2nd order can be applied. But when the process control step is too large, the Adams-Bashforth algorithm could locally generate larger errors. A much better solution would be to apply the Adams-Moulton algorithm correction, which is stable regardless of the process control step.

References

1. Yang, J.M., Kim, J.H.: Sliding Mode Control for Trajectory Tracking of Nonholonomic Wheeled Mobile Robots, IEEE Transaction on Robotics and Automation Vol. 15 No. 3 (1999) 578–587
2. Martins, F.N. Celeste, W.C, Carelli, R., Sarcinelli-Filho, M., Bastos-Filho, T.F.: An adaptive dynamic controller for autonomous mobile robot trajectory tracking, Control Engineering Practice 16 (2008) 1354-1363
3. Giergiel, M.J., Hendzel, Z., Żylski, W.: Modeling and control of wheeled mobile robots, Warsaw, Polish Scientific Publishers PWN (2002)
4. Beniak, R., Pyka, T.: Influence of PWM on Trajectory Accuracy in mobile Robot Motion, JAMRIS Vol 6 (2012) 50-56
5. Chua, L.O. and Lin, P.M.: Computer Aided Analysis of Electronic Circuits: Algorithms and Computational Techniques, Prentice-Hall, Englewood Cliff, New Jersey (1975)

Work co-financed by European Social Fund



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY





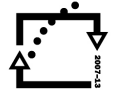
eu
european
social fund in the
czech republic



EUROPEAN UNION



MINISTRY OF EDUCATION,
YOUTH AND SPORTS



OP Education
for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT

NETFEI Workshop

NETFEI workshop – organized as a part of WOFEX 2013 workshop – have been realized as a part of NETFEI project (CZ.1.07/2.4.00/31.0031). This project is co-financed by ESF and Czech state budget.



Acknowledgement

Following papers were supported by the NETFEI – Networks and partnership development between the Faculty of Electrical Engineering and Computer Science VSB-TUO and private sector and universities, reg. no. CZ.1.07/2.4.00/31.0031, co-financed by ESF and state budget of the Czech Republic:

- Maroš Ďurica. *Calculation of Locomotive Tractive Power During the Starting Phase*
- Marcel Fajkus, Jakub Cubik, and Stanislav Kepak. *The convergence of OpenBTS and OpenIMSCore projects*
- Jiří Friedrich. *Sensorless Control of Induction Motor Using State Observers*
- Ondřej Grunt. *Risk Modeling of Accidental Events and Responses in Process Industries*
- Václav Hapla. *Efficient Implementation of the TFETI Coarse Problem Solution in the FLLOP library*
- Martin Hasal. *Comparison of laparoscopic and open surgery techniques in liver surgery*
- Stanislav Hejduk, Jiří Bocheza, Karel Witas, and Artem Ganiyev. *Visible Light Communication with Four Wavelengths Division Access*
- Barbara Helštýnová, Karel Sokanský, and Tomáš Novák. *Comparison of photometric properties of the projection screen and interactive whiteboard with EN 12464-1 at classroom*
- Radim Hercík. *Safe HMI and Car to Car Communication*
- Marek Horinek and Petr Bilik. *Power Analyzer for Measurement of Efficiency of the Frequency Converter*
- Vít Houdek. *Technical condition and importance of line for 110 kV power lines*
- Rostislav Hrtus, Jaroslav Haslinger, and Radim Blaheta. *Identification problems with a priori given material interfaces*
- Jiří Janša. *Annual Energy Balance of Biogas Station*
- Kateřina Janurová. *Weibull Proportional Hazards Regression Model for Patients with Surgery of Colon*
- Jakub Jaros, Andrej Liner, and Martin Papes. *Measurement of homogeneity of the microwave emitter*
- Jaromír Konečný. *E-Bike Testing Station*
- Jiří Krček and Jaroslav Vlček. *Tangential Fields in Mathematical Modelling of Optical Diffraction Problems*
- Jakub Mañas. *Hybrid Power System Power Flow Analysis*
- Oliver Marcinčin and Zdeněk Medvec. *Active charging stations for electric cars*
- Michal Merta. *Parallel implementation of fast boundary element method*
- Martin Mikulec, Jan Rozhon, Filip Řezáč, Jiri Vychodil, and Miroslav Vozňák. *Integration of voice services into corporate information systems*
- Kristina Motyčková and Radek Kučera. *Semi-Smooth Newton Method for Contact Problems with Tresca and Coulomb Friction*
- Jana Nowaková and Miroslav Pokorný. *Knowledge-based Monitoring and Re-adaptation of PID Controllers*
- Vít Otevřel and Zdeněk Slanina. *Remote control of laboratory task using wireless technologies and operating system Android*

- Pavol Partila, Karel Tomala, and Jakub Safarik. *Recognizing emotions from human speech using 2-D neural classifier and influence the selection of input parameters on it's accuracy*
- Lukas Peter and Martin Cerny. *Verification methodology for noninvasive continuous measurement of blood pressure*
- Petr Pětvaldský. *Testing gapless measurement of PQA according to IEC 62586-2*
- Phan Thi Thanh Thao and Lubomír Ivánek. *Study of Electromagnetic Compatibility between Railway Infrastructure and Train Interface*
- Michal Reguľa. *Power quality measurement*
- Aleš Ronovský. *Parallel Hybrid Mesh Multiplication*
- Petr Rozehnal. *Complaints about the quality of electricity in Northern Moravia*
- Stanislav Rusnok, Pavel Sobota, and Pavel Svoboda. *The Use of the General Theory of Electrical Machines for Modeling of Induction Motor in Matlab Simulink*
- Pavel Skalný. *Applied graph theory in Markov chains reliability analysis*
- Roman Slabý and Martin Stankuš. *System for Automatic Loading under Operating System QNX*
- Ladislav Slíva. *Identification and Optimization of Parameters of Synchronous Machines with PM*
- Martin Smoček and Zdeněk Hradílek. *Photovoltaic Power Plant, Extreme Thresholds of Production*
- Maros Smondrk, Michal Vozda, Jana Krohova and Barbora Hrvolova. *The SSVEP Detection Using the Canonical Correlation Analysis and the Continuous Wavelet Transform*
- Martin Stachoň. *Comparison of numerical codes for diagonalization of Hermitian matrices*
- Jindřich Stuchlý. *Monitoring of Hybrid system using LabView – Next step in development*
- Dominik Szabó. *Modelling of a DVR model based on DQ transformation control technique in MATLAB/Simulink*
- Jan Šrámek. *Design, realization and analysis-methods for measurement heat pump energy balance*
- Jan Šumpich, Zbyněk Carbol, and Karel Sokanský. *Measures to achieve energy savings of lighting systems using daylight*
- Lucie Švecová. *RFID Expert Database*
- Marián Uher. *Analysis of power grid including energy storage station using EMTP-ATP*
- Jan Vaculík. *Energy rating of hydrogen system at VSB-TUO*
- Alena Vasatova and Aleš Ronovský. *Elastic image registration with local refinement*
- Tomáš Verner. *Sensorless Control of Induction Motor Using Injection Method*
- Jan Vitásek and Jan Látal. *Recovery of LED Diode Optical Spectrum*
- Petr Vojčínák. *Using Mathematical Background for Determining the Temperature Distributions and Corresponding Parameters inside a Rock Massif*
- Jakub Vramba. *Off-grid hybrid power system*
- Jan Zapletal. *Shape Optimization Based on the Shape Calculus and the Boundary Element Method*

Calculation of Locomotive Tractive Power During the Starting Phase

Maroš Ďurica

Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
d.maros@seznam.cz

Abstract. In railway freight and passenger transport a problem of a locomotive starting uphill with a given train load with regard to adhesive force needs to be solved very often. Magnitude of the starting tractive power of a locomotive depends on a track reduced gradient and a load value, which can be modified by increasing or decreasing a number of wagons. A limiting factor of starting up to speed of a locomotive with a load is an adhesive force which relates to a locomotive weight and a surface of rails.

Keywords. tractive power, adhesive force, reduced gradient, wet and dry rails, locomotive weight

1 Introduction

Adhesion is a physical phenomenon which occurs when transmitting a tractive or braking power by a wheel to a rail. A cause of adhesion occurrence is intermolecular force on an adhesion limit. Adhesion coefficient φ depends on: run speed, railway track superstructure condition, weather (dry – wet), mechanical and electrical arrangement of a vehicle – considering a vehicle weight and a number of driving axles of an undercarriage. When starting up to speed, the adhesion coefficient is less by approximately 70 N/KN while wet than while dry. [1, 2]

Minimum tractive power of a train with a locomotive for starting and braking is calculated as follows:

$$F_{\text{tmin}} = G_L \cdot (p_{oL} + p_{\text{sredmax}} + 102 \cdot \xi_L \cdot \omega_{\text{min}}) + G_v \cdot (p_v + p_{\text{sredmax}} + 102 \cdot \xi_v \cdot \omega_{\text{min}}) \quad (1)$$

(N, kN, N · kN⁻¹, N · kN⁻¹, m · s⁻²)

where G_L , p_{oL} , ξ_L are values applying for a locomotive,

G_v , p_{ov} , ξ_v are values applying for wagons.

For starting up to speed of a locomotive on a given track segment in the most unfavourable conditions the following equation applies:

$$F_{\text{adhwet}} \geq G_L \cdot (p_{oL} + p_{\text{sredmax}} + 102 \cdot \xi_L \cdot \omega_{\text{min}}) + G_v \cdot (p_v + p_{\text{sredmax}} + 102 \cdot \xi_v \cdot \omega_{\text{min}}) \quad (2)$$

(N, kN, N · kN⁻¹, N · kN⁻¹, m · s⁻²)

F_{adhwet} is an adhesive force while rails are wet; it must be equal or higher than a minimum tractive power to start the train run.

p_{sredmax} is the highest reduced gradient in the given track segment.

2 Data of 181 and 380 type locomotives, a Faccs wagon and a track

A calculation of minimum tractive power for starting up to speed of 181 - E. 661 and 380 – 109E type locomotives has been performed with the aid of Mathcad software for the Bohumin – Čadca track segment at the worst adhesion conditions in dependence on a number of loaded Faccs type four-axle freight wagons for bulky goods of 38 m³ capacity. [3, 4, 5]

Fig. 1. Comparison of parameters for the 181 and 380 type locomotives and the Faccs type wagon




181 type locomotive	380 type locomotive
	
<p>Manufacturer: Škoda Plzeň works Factory identification: E 661.2 Year of manufacture: 1961-62 Undercarriage arrangement: Co'Co' Traction system: 3 kV= Rotational mass coefficient: 1.25 Service weight: 120 t Permanent tractive power: 213 kN Maximum tractive power: 345 kN Permanent output: 2 610 kW</p>	<p>Manufacturer: ŠKODA Transportation Factory identification: 109 E Year of manufacture: 2008-2010 Undercarriage arrangement: Bo'Bo' Traction systems: 3 kV=; 15; 25 kV~ Rotational mass coefficient: 1.25 Service weight: 82 t Permanent tractive power: 213 kN Maximum tractive power: 274 kN Permanent output: 6 400 kW</p>
<p>Faccs type four-axle freight wagon</p>	
<p>Manufacturer: VagónkaPoprad Factory identification: Faccs Year of manufacture: 1970 Number of axles: 4 Carriage nominal weight: 27t Weight of Charge: 57t Loading space – capacity: 38 m³ Rotational mass coefficient: 1.25</p>	

Table 1. Data of Bohumín – Čadca No.320 railroad track

Section No.	Station	Distance (km)	Max. Speed (km/h)	Reduced gradient (N/kN)
1	Bohumín	0	40	0.85
2		1.182	110	6.21
3	Dolní Lutyně	4.052	110	4.16
4		7.902	60	2.16
5	Dětmarovice	8.494	100	3.14
6	Karviná hl.n.	15.223	40	3.99
21	Jablunkov- návší	49.924	70	16.22
22		50.544	70	16.32
23		51.094	80	17.32
24	Bocanovice	53.144	70	14.92
25		54.069	70	16.12
26		55.144	80	14.12
27		57.144	70	15.13
28	Mosty u Jabl.	58.722	70	-14.76
29		61.616	70	-12.3
30		62.144	80	-11.6
31	Svrčinovec	65.785	80	-9.8
32	Čadca	68.842	80	0

3 Calculation of Tractive Power During the Starting Phase for 181 and 380 Type Locomotives

Train speed values in particular stages have been substituted into an equation of specific driving resistance of the 181 and 380 type locomotives and the Faccs wagon. We have obtained P_{O181} , P_{O380} and P_{Ovagon} matrixes. Then we perform a calculation of an adhesive force for wet rails F_{maxwet} and for dry rails F_{maxdry} by product of weight of carriages, adhesion coefficient and a constant determining utilization of the locomotive weight. The locomotive weight is a product of the weight and gravitational acceleration. [6]

$$P_{O181} = 3.8 + 0.02 \cdot V1 + 0.0004 \cdot V2 \quad (N \cdot kN^{-1}, km \cdot h^{-1}) \quad (3)$$

$$P_{O380} = 2.8 + 0.00085 \cdot V2 \quad (N \cdot kN^{-1}, km \cdot h^{-1}) \quad (4)$$

$$P_{Ovagon} = 1.3 + 0.00033 \cdot V2 \quad (N \cdot kN^{-1}, km \cdot h^{-1}) \quad (5)$$

$$p_{181,380,vagon} = P_{O181,380,vagon} + psred \quad (N \cdot kN^{-1}) \quad (6)$$

$$G_{L181,L380,vagon} = M_{181,380,vagon} \cdot g \quad (kN) \quad (7)$$

$$F_{\text{maxwet}_{181,380}} = G_{L181,380} \cdot \epsilon_L \cdot \varphi_{\text{adhwet}} \quad (\text{N}, \text{kN}, 1, \text{N} \cdot \text{kN}^{-1}) \quad (8)$$

$$F_{\text{maxdry}_{181,380}} = G_{L181,380} \cdot \epsilon_L \cdot \varphi_{\text{adhdry}} \quad (\text{N}, \text{kN}, 1, \text{N} \cdot \text{kN}^{-1}) \quad (9)$$

where φ_{adhwet} is an adhesion coefficient while wet and φ_{adhdry} is while dry, ϵ_L is a coefficient of utilization of the adhesive force, it reaches a value 0.90 for a locomotive which has all axles driving, $G_{L181,380}$ is an adhesive weight of the 180 and 380 type locomotive.

Necessary minimum tractive power for starting a train with the 181 type locomotive and eleven wagons is as follows:

$$F_{t181} = G_{L181} \cdot (p_{180} + 102 \cdot \xi_L \cdot \omega_{\text{min}}) + G_{11v} \cdot (p_{\text{vagon}} + 102 \cdot \xi_v \cdot \omega_{\text{min}}) \quad (\text{N}) \quad (10)$$

Necessary minimum tractive power for starting a train with the 380 type locomotive and eight wagons is as follows:

$$F_{t380} = G_{L380} \cdot (p_{380} + 102 \cdot \xi_L \cdot \omega_{\text{min}}) + G_{8v} \cdot (p_{\text{vagon}} + 102 \cdot \xi_v \cdot \omega_{\text{min}}) \quad (\text{N}) \quad (11)$$

$$F_{t181} = \begin{pmatrix} 135716 & 202876 & 187856 & 150782 \\ 174903 & 151394 & 156293 & 168968 \\ 176295 & 160087 & 165505 & 166380 \\ 173783 & 143935 & 174515 & 174826 \\ 187921 & 193386 & 242001 & 216097 \\ 257265 & 257995 & 269310 & 247739 \\ 256532 & 245864 & 249278 & 44971 \\ 48338 & 57453 & 70639 & 142429 \end{pmatrix} F_{t380} = \begin{pmatrix} 97963 & 149782 & 138840 & 109376 \\ 128756 & 109384 & 113133 & 124433 \\ 129771 & 116154 & 120850 & 121487 \\ 126880 & 104722 & 127414 & 128700 \\ 137180 & 140748 & 177638 & 157293 \\ 187285 & 187816 & 196473 & 180345 \\ 186751 & 179392 & 181466 & 32625 \\ 35078 & 42132 & 51738 & 104039 \end{pmatrix} \text{N}$$

Now we obtain matrixes p_{181} , p_{380} , p_{vagon} by summation of a specific driving resistance individually for the 181, 380 locomotives and the wagon with the track reduced gradient.

A hauled weight of the locomotives is created by Faccs wagons, the weight of one Faccs wagon is 57 tons.

The 181 type locomotive has been loaded with 11 wagons, the 380 type locomotive has been loaded with 8 wagons, this has been substituted into equations (10) and (11) and it can be seen from the resulting matrixes F_{t181} and F_{t380} that some values of minimum tractive power for the 181 and 380 type locomotives with hauled weight exceed the adhesive force value for wet rails which is 269 310 N for the 181 type locomotive and 196 473 N for the 380 type locomotive.

It is shown in kN in graphs 1 and 2.

4 Conclusion

In this paper the problem of calculation of tractive power for starting up to speed of 181 and 380 type locomotives with regard to an adhesive force in the given Bohumín – Čadca track segment has been solved.

Bad meteorological conditions which had caused wet surface of rails have been taken into consideration.

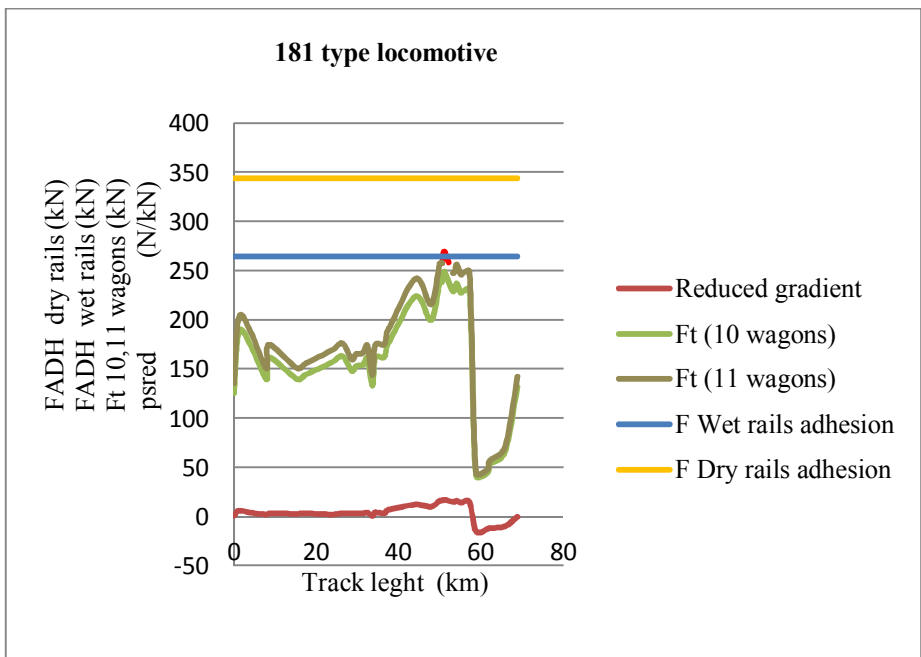
Regarding the relation no. 2, the adhesive force must be higher or equal to the minimum tractive power of a locomotive train for starting up to speed. The most unfavourable conditions for starting are in the track segment no. 23 where the reduced gradient value is 17.32 N/kN.

In this Bohumín – Čadca track segment the minimum tractive power value for the 181 type locomotive exceeds the adhesive force value (wet rails), when 11 loaded wagons are being hauled, as it is depicted by a red dashed line in Graph 1. It can be seen in Graph 2 that for the 380 type locomotive the minimum tractive power is higher than the adhesive force, when 8 loaded wagons are being hauled.

The 181 type locomotive can cope with the given Bohumín – Čadca track when hauling 10 loaded wagons, whereas the 380 type locomotive can haul only 7 wagons in no. 23 track segment, as it is shown in Figure 2.

It is probably caused also by the 181 type locomotive weight which is by a third higher – 120 tons – than the 380 type locomotive weight – 88.2 tons.

Graph 1. Dependence of minimum tractive power on the track reduced gradient for the 181 type locomotive with regard to the adhesive force, while rails are being wet or dry



Graph 2. Dependence of minimum tractive power on the track reduced gradient for the 380 type locomotive with regard to the adhesive force, while rails are being wet or dry

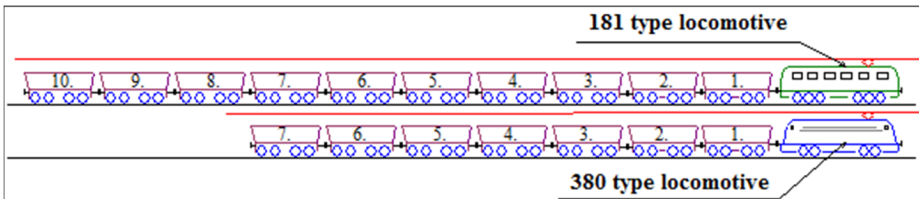
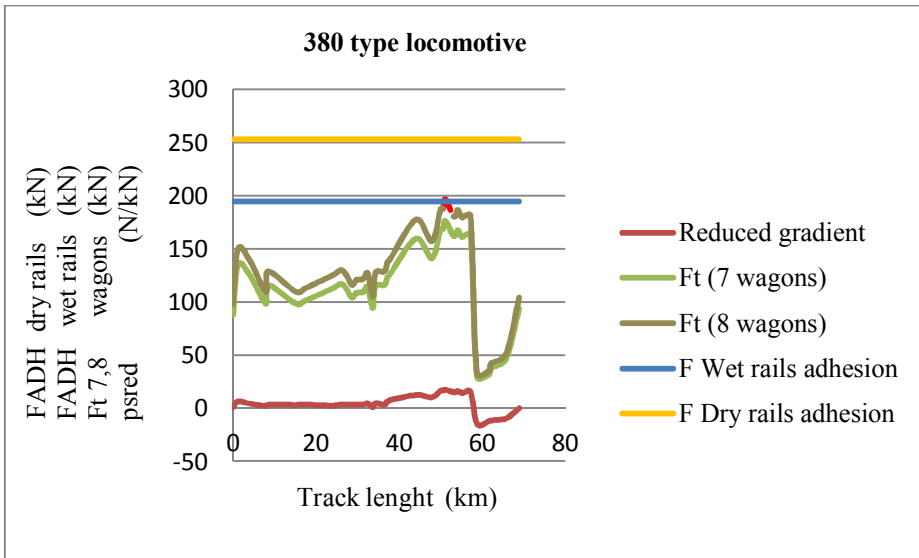


Fig. 2. Rate of a number of loaded wagons to the 181 and 380 type locomotives on the Bohumin – Čadca track

References

1. Paleček, J.: Electroenergetics in Transportation, study materials for lectures, VŠB – Technical University of Ostrava, Winter Semester 2010.
2. Horák, K.: Electric Traction and Energetics in Transportation, VŠD Žilina, 1971.
3. Wikipedia. *Locomotive 181* [online]. c2009. [retrieved. 2013-07-14]. URL: <http://cs.wikipedia.org/wiki/Lokomotiva_181/>.
4. Wikipedia. *Locomotive 380* [online]. c2008. [retrieved. 2013-07-16]. URL: <http://cs.wikipedia.org/wiki/Lokomotiva_380/>.
5. Company RYKO PLUS. *Wagon type Faccs 38 m³* [online]. c2010. [retrieved. 2013-02-01]. URL: <<http://www.rykoplus.cz/doc/typovy-list-faccs-38-m3.pdf/>>.
6. Drábek, J.: (Dynamics and Energetics of Electric Traction), (Alfa, technical and economical literature publishing), Bratislava 1987.

Comparison of photometric properties of the projection screen and interactive whiteboard with EN 12464-1 at classroom

Barbara Helštýnová, Karel Sokanský, and Tomáš Novák

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{barbara.helstynova.st, karel.sokansky, tomas.novak}@vsb.cz

Abstract. Today they are used blackboards on which are written by chalk or marker and the other type is the interactive white-board. Interactive whiteboards can replace both classic blackboard and projection screen. Information displayed by these means, we can see mainly through the eye, therefore we accept visual information. The human eye responds primarily on luminance contrast. The goal of this work is to compare the photometric properties of the projection screen and interactive whiteboard with EN 12464 - 1. Properties are compared under different lighting conditions, in order to give an optimal set of lighting systems.

1 Introduction

Interactive whiteboards have become a new teaching tool not only in universities, but it can already be found even at primary and secondary schools. For the general definition of what is an interactive whiteboard can be considered: “The interactive whiteboard is a touch-sensitive surface, through which runs for active communication between the user and the computer to ensure the maximum possible level of clarity displayed content.”.

The luminance analysis was compared to the interactive whiteboard and projection screen. Interactive whiteboard creates a mirror glare, while the screen can be considered Lambert surface. Measuring the luminance was under different lighting conditions.

2 Description of measurement

For comparison, projection screen and interactive whiteboard classroom was chosen for VSB-TU Ostrava. This work was inspiring because of complaints of glare, not only in the classroom. Screen is the hung above the interactive whiteboard and can be rolled up to the need for screening. In the classroom are surface mounted in two rows of three. Fluorescent lamp is a type 4×58 W with opal diffuser. Each light can be switched on and off separately. The room can

eliminate daylight with shutters. On projection screen and board is projected by the data projector BenQMP626 in preset mode “Dynamic”.

In the classroom were measured using a luminance analyzer these situations while the video player:

1. Fully lighted classroom.
2. The first series of lights off.
3. All lights off.

Each situation was measured at eye seated person (about 1.2 meters) from the first and second rows of benches in three positions, and in the place closest to the door, center – a direct view on the interactive whiteboard / projection screen and place closest to the window (Fig. 1). Was selected as the background color white, which is the worst possible option in terms of the highest luminance, the screen was placed black point that was always focused to the same point of measurement points. Measurement was carried out using the luminance analyzer LMK-based SLR Canon EOS 350D.

The processing of the measured images used software LMK 2000. As a basis for the evaluation of the data used in CR2 format in which the image is recorded in the form of RAW image.

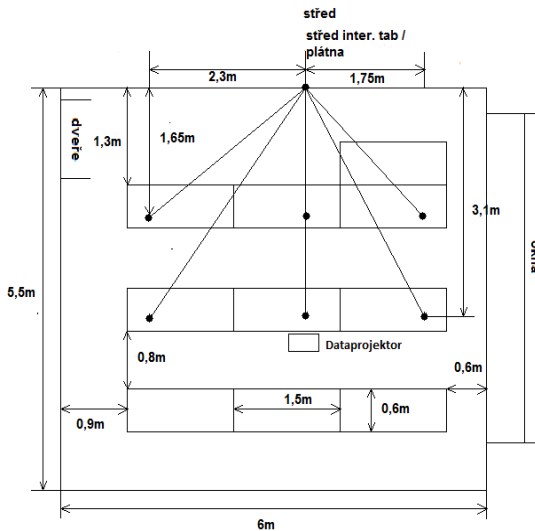


Fig. 1: Sketch plan view classroom

2.1 Projection green

Projection screen is hung on the wall and consists of a white surface with black borders on the left and right side (3 cm). Screen is a pop-up type with dimensions of 1.8×1.8 m.

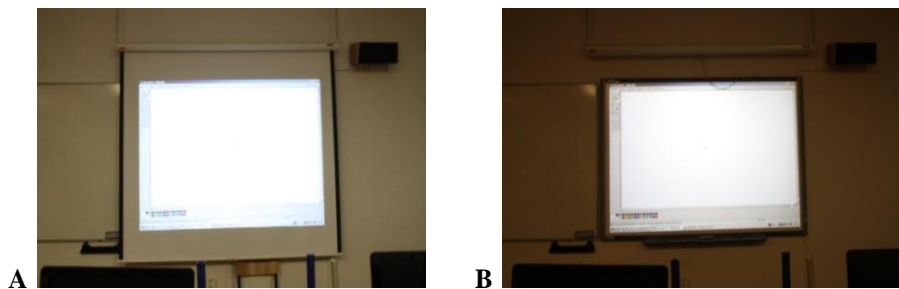


Fig. 2: A - the projection screen, B - the interactive whiteboard

2.2 The interactive whiteboard (SmartBoard Model 680)

Surface of interactive whiteboard is resistive type. Touch board resolution is approximately 4000×4000 pixels. Active area has dimensions of 1.56×1.17 m. Blackboard can be controlled using a PC, you can use a special pen (4 colors-red, blue, green, black) or use the hand as a means of control.

In the classroom interactive whiteboard is used with front projection. The disadvantages include a front projection screen by itself, if it enters into the light beams transmitted data projector. When using a front projection was detected mirror reflection.

3 The norm EN 12464-1

Norm EN 12464-1 specifies the conditions for illuminance of internal work space. The conditions for illuminance are given help instead of visual task and the immediate surroundings of visual task and background visual task (Table 1). Figure 3 shows the defining point of visual task, the immediate surroundings of visual task a background visual task.

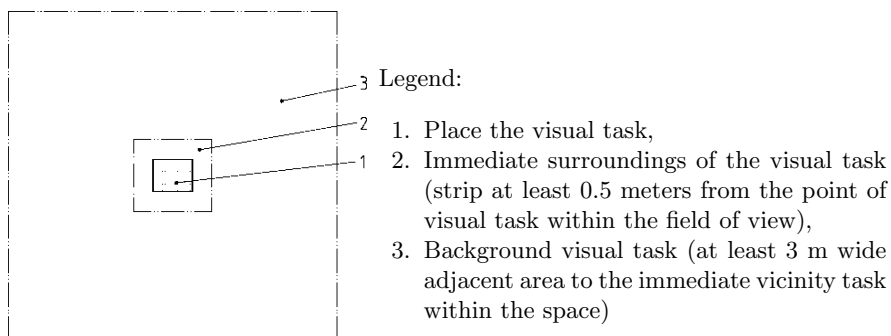


Table 1: Relationship between illuminance immediate environment and task lighting in place of the visual task (EN 12464-1)

Illuminance place visual task E_{task} (lx)	Immediate vicinity of the task illuminance (lx)
≥ 750	500
500	300
300	200
200	150
150	E_{task} (lx)
100	E_{task} (lx)
≤ 50	E_{task} (lx)

4 Illuminance

Vertical illuminance interactive whiteboard and projection screen was measured at the visual task and the immediate surroundings of visual task (distance 0.5 m) using luxmeter.

Table 2: A – vertical illuminance values for in.whiteb. for the disabled lamps above (1st row lights off) – White projection, B – values of vertical illuminance for in.whiteb. for full illumination (all lights on) – White projection

A	0,5m [lux]	Left side [lux]	Middle [lux]	Right side [lux]	0,5m [lux]
1.row	90	860	1100	760	80
2.row	90	890	1200	840	90
3.row	100	780	1090	700	95
E_{AVG} (int.board./screen)		913			
E_{AVG} (0,5m)		91			

B	0,5m [lux]	Left side [lux]	Middle [lux]	Right side [lux]	0,5m [lux]
1 row	450	1100	1380	1060	380
2 row	430	1210	1610	1100	370
3 row	420	1000	1290	1020	330
E_{AVG} (int.board./screen)		1197			
E_{AVG} (0,5m)		397			

According to (Table 1), the value of the visual task illuminance should be higher than 750 lux, in the case of the interactive whiteboard and screen is average 913lux, which means that the illuminance the visual space corresponds to the standard task. The immediate surroundings of the visual task should have the illuminance of 500 lux, but the value of interactive whiteboard cannot be achieved full explanation classroom. The value of illuminance immediate vicinity

of the visual task with full lighting 397 lux and when the condition first of all lights off (directly above the board) value is just 91 lux illuminance in diameter.

5 Luminance analyses

Luminance analysis was performed for the three light modes, as mentioned above. For each measurement point was done 5 RAW image with different exposure time. The resulting image is then composed of those shots with different exposures. To simplify the sensing location, was created on a white background point on which all images have been focused.

When comparing the mean values on the screen, you can see almost the same value for all three measurement locations. On the contrary, the value of interactive whiteboards has the same average value and direct view even vary by up to 100 cd.m². Comparing the maximum luminance values from direct view (Table 3a and Table 3b) is up to 5-fold difference.

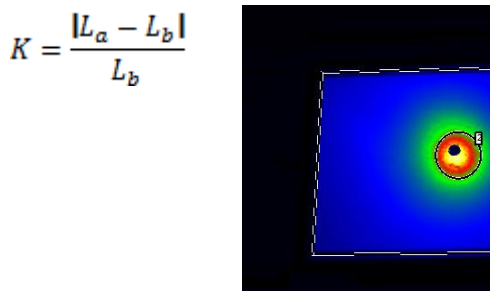


Fig. 3: Luminance analysis of inter. Whiteb., direct view light of the board

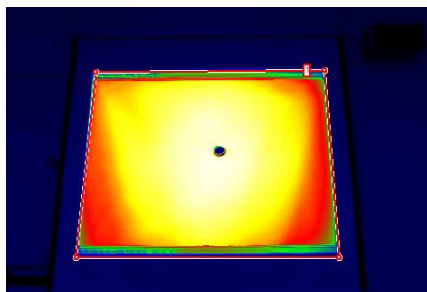


Fig. 4: Luminance analysis of projection screen, direct view – lights off of the green

Table 3: Luminance analysis

(a)			(b)		
	L_{avg} [cd.m ⁻²]	L_{max} [cd.m ⁻²]		L_{avg} [cd.m ⁻²]	L_{max} [cd.m ⁻²]
Direct view	309,7	1490	Direct view	242,1	322,9
Mirror reflection	902,7	1490	Mirror reflection	---	---

5.1 Luminance distribution

Change Z1 EN 12464-1 from 2005 shows the optimal ratio of the brightness of visual space mission to the brightness of the immediate vicinity of the visual task (maximum distance of 0.5 m) and the brightness of the background (distance 3 m – field of view of the observer) at a ratio of 10 : 4 : 3

Table 4: Ratio of luminance interactive whiteboards and screens compared with the norm

Ratio	10	4	3
Direct view	294,5	129,3	19,36
Pr. screen	10	4,4	0,66
Direct view in. board.	934,3	177,2	18,35
	10	1,89	0,19

5.2 Luminance contrast

For the resolution of the observed objects in the visual field, it needs to be sufficiently different subjects' luminances. Luminance contrast by which we judge the visibility of objects is defined by the relation: K – contrast, L_a – observed object (mirror reflection, [cd.m⁻²]), L_b – surroundings (interactive whiteboard without mirror reflection [cd.m⁻²]).

6 Conclusion

When analyzing the luminance interactive whiteboard was detected mirror reflection from all points of measurement (example Fig. 3). Mirror reflection can be observed even when the data projector, including off the board. In this case, the mirror surface mounted lamps. The canvas mirror reflection arises; on the contrary, we can say that the screen is a Lambert surface with diffuse reflection.

Table 5: Luminance contrast interactive whiteboard

View door	L_a [cd m ⁻²]	L_b [cd m ⁻²]
	448,7	163,2
Contrast	1,75	
View center	L_a [cd.m ⁻²]	L_b [cd.m ⁻²]
	902,7	261,5
Contrast	2,45	
View window	L_a [cd.m ⁻²]	L_b [cd m ⁻²]
	609,3	186,7
Contrast	2,26	

This statement is confirmed by the average brightness values that are around the value 242 cd.m².

The luminance distribution that shows a change in standards EN 12464-1 Z1 should be in the ratio 10: 4: 3 to place the visual task, the immediate surroundings of visual task and the background. The interactive whiteboard is the ratio of nearly 10: 1.89: 0.19 (Table 5), which is not the standard. However, the standards screen also does not answer, the ratio is 10 : 4.4: (Table 5). To improve the brightness ratio, it would be necessary to increase the brightness of the surroundings and reduce the brightness of the background or interactive whiteboard and screen. It would also be suitable for optimum power lighting system, change the type of interactive whiteboard's surface, matt and glossy for possibly change the data projector distance from the board and its slope.

References

1. Škoda, J., Baxant, P., Fotometrie svítidel s využitím digitální fotografie, SVETLO 2009, Zborník z 18. medzinárodnej konferencie;
2. Helštýnová B., Sokanský K., Novák T., Luminance analyse of the interactive whiteboard with Compare to projection screen. Conference "EPE 2013", Kouty nad Desnou
3. ČSN EN 12464-1. Světlo a osvětlení – Osvětlení pracovních prostorů – Část 1: Vnitřní pracovní prostory. Praha, 2012. 56 p.
4. ČSN EN 12464-1 ZMĚNA Z1. Světlo a osvětlení – Osvětlení pracovních prostorů – Část 1: Vnitřní pracovní prostory. Praha, 2005. 12 p.

Technical condition and importance of line for 110 kV power lines

Vít Houdek

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
vit.houdek@centrum.cz

Abstract. The article deals with database of faults, measured data and backup alternatives for 110 kV power lines. This is important data to determine the technical condition and importance of line. Technical condition and importance of line are input parameters for restoration method for 110 kV power lines. Restoration method is based on reliability centered maintenance (RCM). The needed data are: faults on power lines which provide information on technical condition, amount of distributed energy and effects of a failure to show importance of a line.

Keyword: RCM, power line, technical condition, importance of line, restoration method

1 Introduction

Reliability centered maintenance is generally applied on different elements in the same way. Firstly, importance and technical condition of a given element must be assessed. The assessment of these parameters differs from an element to element.

In case of 110 kV power lines, the technical condition is assessed from the failure database of the particular distribution area. The importance of a power line is assessed on the basis of two parameters: distributed power and the effect of a line failure. To our benefit, all the mentioned parameters are monitored and recorded in some way by the distribution network operators. The above mentioned parameters shall be elaborated in the following paragraphs.

The aim is to develop a methodology for restoration 110 kV power lines, which depends on technical condition and importance of line.[1]

2 Technical condition

The database of faults on 110 kV power lines gives information on the technical condition. The data is from 5 distribution areas. The system of recording has changed several times in history, but since 1.7.2009 a mandatory integrated system has been established. First records are from 2002 – Table 1 shows numbers of records and time periods. So far there are three record periods: to 2006, from 2007 to 6.2009 and from 1.7.2009 (the table shows fault up to 26.9.2012). This distribution of records is due to major difference in recording of faults into the database.

Table 1. Count of records in database of faults

Time period	to 2006	from 2007 to 6.2009	from 7.2009	All
Distribution area	Number	Number	Number	Number
DA 1	2228	1633	2438	6299
DA 2	0	635	1421	2056
DA 3	173	1187	2064	3424
DA 4	7	242	1274	1523
DA 5	0	80	1822	1902
DA 1 - 5	2408	3777	9019	15204

The distribution companies need a database which can record various pieces of information on a fault. The distributor’s internal system gives a database in a format suitable for EXCEL. Altogether, 213 pieces of information on fault can be found, e.g. serviceman’s name, cause, priority, quantity etc. However, most of this data is irrelevant and therefore removed from final assessment. The most important are: priority, quantity, maintenance area, distribution area, and code (type) of fault. Priority of fault tells when it is necessary to remove the fault. There are four priorities of faults in the database, today (Table 3). Based on the priorities of fault was made basic evaluation.

An important parameter – quantity – is ambiguous. When the fault line or earth wire, the quantity is given in meters – in other words the entry is the length of the part in which there is a failure (e.g. tree branches in the power lines). On the other hand, the amount means a number of items (pieces) if the fault is on warning notices or isolators. After independent evaluation of identical items, three categories for quantity were determined. (Table 2)

Table 2. Categories for quantity

Quantity - category			
Meters		Pieces	
to 25 m	1	to 10 pcs	1
from 25 to 50 m	2	from 10 to 30 pcs	2
above 50 m	3	above 30 pcs	3

Concerning faults, it must be noted that they are not recorded with respect to a particular line but a technical spot. Technical spot is a section of a line – in other words its maintenance section. Some lines have one maintenance section, some have more of them. In total, there are 388 maintenance sections in the database. [2]

Table 3. Evaluation of priority of faults with quantity

Priority of Fault	Quantity		
	1	2	3
1 - Urgent	127	1	
3 - Month	78		3
4 - Year	688	63	30
5 – Control at ŘPÚ	327	65	39

2.1 Determination of technical condition

Concerning technical condition, only quantity is categorized. There can be more individual records with only one priority in one maintenance section, therefore we need to know the total number of failures with the same priority for the same quantity category. This can be easily determined from the 110 kV power lines failure database for each maintenance section. We use contingency table of MS EXCEL for the determination of failure quantity.

To directly determine the technical condition TS after the evaluation of quantity, priority and maintenance section would be very complicated concerning the development of future software. Therefore a ‘point technical condition’ BTS was introduced. This can be calculated in the following equation (1):

$$BTS = \sum VP_x \cdot \sum (p_{M1x} + k_2 \cdot p_{M2x} + k_3 \cdot p_{M3x}), \tag{1}$$

where VPx is the value of the weight for priority number x. Further, p_{M1x} is the number of failures for priority number x and quantity category 1. Similarly, the number is the same for other quantity categories. Coefficients k₂ and k₃ help to increase the point evaluation for higher quantity categories.

From total value the TS technical condition can be determined. The initial method design counts with five categories (Table 4). To debug the method, a functional file for MS EXCEL was created, whereby all weight values and coefficients can be modified. Hereby given values are not to be considered terminal. They are meant to explain the whole system for the evaluation of the technical condition of power lines which will enter the future software for the restoration of 110 kV power lines on the principles of RCM.

Table 4. Evaluation of priority of faults with quantity

Value ranking for BTS				
50	100	200	400	>400
Technical condition TS (%)				
95%	80%	70%	60%	40%

3 Importance of line

Assessment of line importance was theorized about in a previously published paper [3]. There, back-up alternatives and their ranking were designed with the method of reliability schemes. This theoretical analysis gave us a short overview of 110 kV power lines back-up alternatives and has been used to evaluate one distribution area.

3.1 Effect of failure

The term “backup alternative” has been replaced by the term ”effect of failure”, as the latter directly denotes the consequence which takes place if a 110 kV power line fails, meaning what the type of its backup is. The distribution field experts helped us to evaluate a hundred and sixty-eight 110 kV power lines in total from one distribution area. For each failed line there is one effect of failure to be selected out of seven alternatives. The evaluation result is that almost 56 % of lines are backed up by a hot reserve, i.e. no manipulation in the distribution network is necessary. Further results are in Table 5.

Table 5. Effect of failure

Effect of failure	Count (-)	Count (%)	Name of effect
1	93	55,95%	“No effect” – no manipulation necessary
2	52	30,36%	Back-up by 110 kV line of approx. the same length
3	1	0,60%	Back-up by 110 kV line of approx. the same length - limited power
4	12	7,14%	Back-up by a longer 110 kV line
5	1	0,60%	Back-up by a longer 110 kV line - limited power
6	9	5,36%	Back-up by MV line
7	0	0,00%	Back-up by MV line - limited power

3.2 Distributed power

The assessment of 110 kV power lines importance on the basis of solely on the effect of failure would be incomplete without counting in the power distributed by the line. This parameter is easy to determine from the power outputs measured in a distribution area. To calculate the distributed power we used data measured in 2010 which were at our disposal. These are 365 XLS files of 20.1 MB each. A file of a day record comprises measured active power, voltage, current, reactive power and other

system data. The values are recorded hourly and their average, maximum and minimum levels are given. We used the average values. The daily record contains measurement on 110 kV power lines plus the measurements on 22 kV lines, substation taps, transformers etc.

The calculation of total energy had several steps. First, total energy in one day was calculated and saved in a new file. The energy from the following day was recorded in a column next to the first one in the same file and the procedure was repeated on and on. In order to check the numbers, the evaluation was divided into two terms which were subsequently added. All these operations had been programmed as macro in the MS EXCEL table processor.

When evaluating distributed energy, it is very important to verify the accuracy of calculated numbers. Abnormalities occurred during the calculation of energy for particular lines. Gradually, a control system was devised using functions in MS EXCEL (COUNTIF, IF a COUNT).

3.3 Determination of line importance

For the future design of RCM software it is not important to know the exact value of distributed energy. Four categories of the energy distributed in a year were selected. The results are in (Tab.6). Distributed energy data were added to the evaluation of the effect of failure. Altogether, there are 28 possible combinations of the effect of failure and distributed energy.

Table 6. Evaluation of distributed power

Distributed power		Count (-)	Count (%)
Value range	Category		
to 50 GWh	1	47	29,01 %
from 50 to 150 GWh	2	61	37,65 %
from 150 to 250 GWh	3	34	20,99 %
above 250 GWh	4	20	12,35 %

As mentioned above, the final importance of a line is determined from the distributed energy and effect of a failure. With these two pieces of input information - both of them categorized – we can use a matrix in Table 7.

The values in the matrix are selected on the basis of previous analyses of input importance and theoretical analysis. It is evident that the power line backups which transmit reduced power have considerably higher values. Also, MV line backups have higher values compared with HV line backups. The same holds for the length of the line backup. Concerning distributed energy, the higher the value of distributed energy, the higher the final value of line importance.

Table 7. Importance of line - matrix

Matrix Importance of line		Distributed power			
		1	2	3	4
Effect of failure	1	0%	5%	10%	20%
	2	5%	10%	15%	25%
	3	15%	25%	30%	45%
	4	7%	12%	17%	27%
	5	25%	35%	40%	50%
	6	20%	25%	30%	40%
	7	35%	45%	55%	70%

4 Conclusion

The introduction of this paper describes the issues of the restoration of 110 kV power lines on the principle of RCM. In second and third section are deal with input parameters to restoration method – technical condition and importance of line. Evaluations of both parameters are described in these parts of the article. This section also describes the problems that occurred during the evaluation of technical condition and distributed energy.

Our further aim is to accurately analyze our evaluation of input parameters and then we have to use them to calculate priority of restoration. Create a functional file for MS EXCEL to analyze and design type power lines.

ACKNOWLEDGEMENTS

This work was supported by the Czech Science Foundation (No. 102/09/1842), by the Grant of SGS VŠB - Technical University of Ostrava (No. SP2013/137) and by the project ENET (No. CZ.1.05/2.1.00/03.0069).

The results of the thesis are published in:

- konference CIRED 2012, Tábor, ISBN 978-80-905014-1-6
- konference ELNET 2012, Ostrava, ISBN 978-80-248-2916-5
- konference EPE 2013, Kouty nad Desnou, 978-80-248-2988-3
- článek v AEEE 3/2013, ISSN 1804-3119

References

1. Moubray J.: Reliability-centred Maintenance. Butterworth-Heinemann, 1997
2. Skog J.: Maintenance Task Interval Determination. Maintenance and Test Engineering Co. USA, 1999
3. Houdek, V., Sojka, J., Rusek, S., Goňo, R.: Backup Alternative for 110 kV lines. EPE 2012. ISBN 978-80-214-4514-7

Annual Energy Balance of Biogas Station

Jiří Janša

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jiri.jansa.st@vsb.cz

Abstract. This paper describes their own consumption, biogas technology, as well as flows of energy in cogeneration units. Chapter deals with the possibility of using biogas as a storage resource for improving the quality of supply in the local electricity distribution system. In conclusion discusses the annual energy balance of agricultural biogas plants.

1 Introduction

Being one of a few renewable resources of electric power, biogas stations are easily controllable and predictable from performance point of view. Their output can be controlled by remote means, which makes them one of the potential storage resources using biomass as the primary source of energy. However, the performance generated by these resources is fairly low rendering them unsuitable for higher voltage levels. In spite of that, they are mostly situated in the countryside and close to power line ends and very convenient as subsidiary supply units for local distribution grids that have become more prone to quality of electric power due to connection of a large number of photovoltaic power plants within.

1.1 Own Electric Power Consumption

The electric energy consumed by a BGS is very hard to enumerate in accurate figures, as the volumes differ per technology. Yet it is still possible to identify certain main components of technologies associated with substantial electric power consumption within. The largest consumers of electric power can be found among appliances used to process the input material. Those are heavy electric power consumers as crushers, for example. Further heavy consumers include a pump/pumps for the integrated pumping system of BGS. Another aspect not to be missed is consumption generated by the fermenter contents mixing unit - units, which depends on the fermentation tank size. Further appliances comprise the very CHPU's (pumps integrated in motor, motor electronics, gas vent) and potential devices for processing of digested product (centrifuge). These are also components with their consumption substantially different from normal values and it can indicate insufficient engineering solution.

The total operating consumption of electric power may even range in the area of approx. 15–30 % of the energy produced, depending on the unit size and technology installed. [2]

1.2 Own Heat Consumption

Utilisation of heat is strongly affected by the ingredients in input raw materials. Genuine agricultural BGS would project the own heat consumption figure far below levels shown by BGS for processing of waste. The law implies that biologically degradable waste must be deprived of pathogenic germs. This process comprises heating the input raw materials to a high temperature required to kill all hazardous microorganisms. That may be even up to 80°C for a period exceeding 1 hour. This process needs to be applied especially on kitchen and slaughterhouse wastes. The remaining heat consumed is used to keep the temperature inside fermenter stable. It depends on the quality of fermenter heat insulation and obviously the actual season of the year, rising several times in winter compared to figures achieved during summer months. That is the reason for very broad range of own heat consumption of 30 % to 80 % of energy produced. [2]

2 Co-Generation Unit

The co-generation units currently used within biogas stations are mostly represented by diesel combustions engines modified for biogas combustion purposes. This modification lies in replacement of the fuel system. There are two methods of fuel ignition used, with adding of diesel into fuel to ensure its ignition upon achievement of certain compression and resultant heat-up of fuel to the ignition point. The disadvantage of this system is large consumption of diesel during higher performance operation, when the diesel content might exceed 10 % of the fuel volume. Another option deals with fitting a glow plug in the motor associated with certain technical problems as these motors are not designed for such solutions.

The motor is coupled with generator via a clutch. The electric efficiency of co-generation unit ranges around 30 to 40 % of energy contained in fuel. *Together with heat efficiency reaching values of 45 to 60 %, the total efficiency of CHPU amounts to approximately 90 %_[1].* Current CHPU designs are being supplemented with the organic Rankin-Clausius cycle (ORC) represented by a modification of power plant Rankin-Clausius cycle with the only different that water vapour driving the turbine is replaced with organic liquid featuring a lower point of volatilisation (70-130°C), while operating pressure is preserved. When using the ORC, flue exchanger is coupled with steam turbines connected to the generator instead of heat off-take. That will increase the electric efficiency up to 45 % (sometimes even more), while the heat efficiency will be reduced at the same time. *The heat output of CHPU require to operate the ORC is equal to approx. 10 kW_i/kW_e of the ORC performance_[4].*

3 BGS Utilisations Options

3.1 BGS as Backup Compensation Supply

One of the options to utilise controlled BGS is to connect it on the same line as a non-controlled, stochastic power supply, e.g. photovoltaic or wind power plant. The BGS serves as a compensation supply in this case and it produces electric power and heat during periods of reduced supply from the primary source only. The specific area would then experience reduction of peak output and achievement of better production stability. As far as their control is concerned, both supply units would have to cooperate together and the BGS would require a larger gas container. For up to half a day production of biogas and with double output motor. It is more convenient rather for collaboration with a photovoltaic power plant that ensures better operation stability compared to wind power plants, which might incur several days of downtime period due to zero wind. [3]

3.2 BGS as Supporting Services

Potential connection of BGS as a supporting service will require improvement of its storage capacity for produced biogas. Six hours of downtime will require raising the motor output by 30 %. Stopping, launching and reduction of production have been addressed at existing BGS facilities for standard conditions of power supply connection and disconnection under crisis situations. The existing control software of BGS has been already modified accordingly and enables full remote control of whole BGS operation. As BGS units feature a relatively low output (regularly up to 1MWe of installed output) and they are greatly dispersed mainly over countryside areas, their significance lies mainly in provision of supporting services to operators of distribution grids. Owing to the increase in production of electric power by photovoltaic power plants, this method for BGS utilisation represents one of the options for maintenance of required electric power quality. This method will be even better to exploit following introduction of the so called "Smart Grids". Provision of supporting services to the transfer electric power grid requires joining of several BGS units into virtual blocks with the minimum output of 10 MW and up to 30 MW in optimal case. The variety of owners possessing individual BGS units poses a problem to utilisation of confined control output only. The reason is, whether all BGS's reduce their output evenly or in proportion to their nominal output only. Higher efficiency will be preserved by shut-down of one or more BGS units. That will invoke certain problems associated with distribution of profits. Another disadvantage would be the continuous production of biogas running independent of CHPU operation. That implies the need to provide such services in regular (daily) intervals to avoid non-productive combustion of stored biogas by the safety burner. [3]

3.3 Control Using Natural Gas Distribution Network

The natural gas distribution network can be used for distribution of biogas process to match the quality of natural gas. That will help towards achievement of maximum utilisation of energy potential from the obtained fuel. The storage effect of distribution network serves mainly to correct the imbalance between almost stable production of biogas and peak operation of the co-generation unit. This method is not utilised to a greater extent due to significant costs incurred by processing of biogas. It is considered for large power supply units in future. However, further progress of technologies applied for biogas processing might result in reduction of investment amounts required for the technology to such extent that the solution could pay its way for smaller power supply units as well. [3]

3.4 Involvement of BGS in Electric Power Equalizing Market

The law in energetics stated that every BGS operator must be registered for this particular market. Each registered participant then has the opportunity to participate in the equalizing market and offer its control capacity within short-term electric power trading transactions. Those are simple transactions between producers and the power grid operator. As it is mostly negative control energy, this is a good opportunity for BGS. The disadvantage of overall shut-down is mainly the increased heat load on certain parts resulting in substantial reduction of their service life. It is therefore better to reduce the output to approximately 30 % of the installed output. Even though the minimum output is not limited, the BGS would be more convenient when connected into a virtual block with higher output. [3]

4 Farming Biogas Station

The biogas station is situated in the premises of farming enterprise concerned with pork farming, besides other. The input raw materials for production of biogas include maize silage and pig's slurry, which represents a sufficient source of liquid for the implemented wet fermentation technology. One of the reasons for this layout was to utilise the exhaust heat for heating in adjacent pigsties.

The installed output of BGS is 1,090 kWe and its calorific power is 1,080 kWt. The initial set-up included 3 co-generating units (CGU) only; those were two diesel units with output of 2×250 kWe and one petrol unit with output of 340 kWe. Due to excess in biogas production, these units were further supplemented with another diesel with identical output of 250 kWe.

The electric power is fed through a transformer station rated 0.4/22 kV into the outside line operated by a power distributor. The transformer station comprises one transformer unit with the maximum output of 1,250 kVA and the nominal current of 1,804 A.

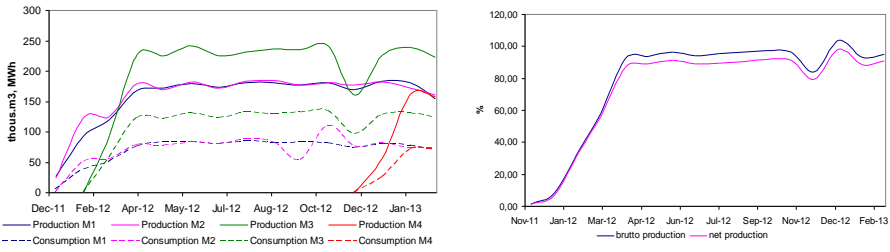


Fig. 1. Course of biogas consumption and electric power production at individual co-generation units

Fig. 2. Course of terminal production and supply of electric power into the electric power distribution network (utilisation percentage of BGS)

The chart 1 shows a gradual approach BPS since December 2011, when the first two were launched KJ, each with an output of 250 kW. Approximately two months later was desolate third cogeneration unit with an output of 340 kW. The fourth KJ of power was put into operation before the end of 2012. Due to the necessary technical adjustments in the engine room a decrease in monthly consumption and production for the third KJ, this had to be shut down for a few days. The graph shows almost constant performance from all cogeneration units with a maximum deviation $\pm 2.5\%$.

The chart 2 shows that the terminal electricity production is holding steady at around 96% of installed capacity with a maximum deviation of $\pm 2.5\%$, indicating good technological discipline. It is also seen that the consumption of the stable at the 5%, which is excellent value.

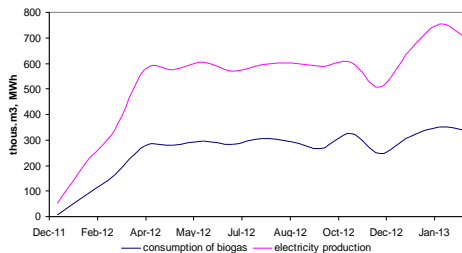


Fig. 3. Course of biogas consumption and electric power production for the whole biogas station

The chart 3 shows how production is directly related to the actual production of biogas and almost perfectly replicates the shape of the curve.

5 Conclusion

From this article you can clearly conclude that the current biogas plants are not equipped for gas storage for longer than a few tens of minutes. Therefore, their performance depends on the actual production of biogas and power fluctuations accurately traced the curve of biogas production. The possibility of using biogas as compensatory resources or support services would be necessary to build biogas tanks for at least 3-6 hourly production and at the same time increase the performance of cogeneration units.

Acknowledgement

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (No. SP2013/137), by the Czech Science Foundation (No. GAČR 102/09/1842) and by the project ENET (Research and Development for Innovations Operational Programme No. CZ.1.05/2.1.00/03.0069).

References

1. **KOGENERAČNÍ JEDNOTKY TEDOM – PŘEHLED VÝROBKŮ** [online]. [cit. 2013-01-15]. Dostupné z: [HTTP://KOGENERACE.TEDOM.COM/KOGENERACNI-JEDNOTKY-DOWNLOAD.HTML](http://kogene.race.tedom.com/kogene.racni-jednotky-download.html)
2. Brandejsová E., Příbyla Z. **BIOPLYNOVÉ STANICE (ZÁSADY ZŘIZOVÁNÍ A PROVOZU PLYNNÉHO HOSPODÁŘSTVÍ)** [CD], 2. aktualizované vydání, Praha, Gas, 2010
3. Matějka J., Štambaský J. **OPTIMALIZACE A REGULACE OZE** [online]. studie 2010, [cit. 2012-12-06]. [DOSTUPNÉ Z: HTTP://WWW.CZBA.CZ/PROJEKTY/OPTIMALIZACE-A-REGULACE-OZE.HTML](http://www.czba.cz/projekty/optimalizace-a-regulace-oze.html)
4. Šafařík M. a kol. **VYUŽITÍ TEPLA Z KOGENERAČNÍ JEDNOTKY**. Zemědělec. 2012, č.21. Profí Press s.r.o.

Active charging stations for electric cars

Oliver Marcinčin and Zdeněk Medvec

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{oliver.marcincin.st, zdenek.medvec}@vsb.cz

Abstract. This paper is about active charging stations for electric cars. Active charging station is usually part of modern electrical grids, known as a Smart Grids or Micro Grids. This paper also include executive summary existing knowledge about the structure and service of electricity distribution network, analysis of electric car industry and its potential impact on electrical network and also conceptual solution charging stations electric cars with minimal effects on power network.

1 Introduction

Electromobility is currently perhaps the most commonly discussed terms of automotive technology. Low number of vehicles operated in the Czech Republic is also corresponding a low density of charging stations, most of which are also operation of an amateur way. In order to be declared the expansion of electric vehicles will increase the demands on the construction of an adequate network of charging stations.

For these purposes seems to be very advantageous to combine the emerging system of infrastructure charging stations with renewable power sources such as the sun on the possibility of accumulation of energy produced and its subsequent delivery to the uniform charging infrastructure.

2 Connecting charging stations for electric cars to the distribution network

If the number of operators of electric vehicles will increase, will be necessary to build sufficiently dense network of charging stations available to ensure not only the range of electric cars in its area of operation, but also allows for the occasional crossing at medium distances up to about 300 km. This requires the deployment of charging sta-

tions along major roads within a maximum of 50-60 km. Their density is present in the Czech Republic is very low (Fig. 1.) and is essentially unmatched density in the neighboring Germany (Fig. 2).



Fig. 1. Map of electric charging stations in the Czech Republic [1]



Fig. 2. Map of electric charging stations on the territory of Germany [1]

2.1 Estimation of the amount of electrical energy needed for charging vehicles

The amount of electric power consumed can be estimated based on the expected performance vehicle structures and their fail safe distances. The need for energy average range electric vehicle 120 km in three distinctive categories gives us tab. 1. For the analysis we selected three basic categories of electric vehicles [2] characterized by different engine power. Category "A" - a small urban vehicle traction engine power 10 kW, category "B" - standard vehicle lower middle class performance traction motor of about 30 kW and category "C" - a higher class of 80 kW. Range will be in all three categories averaged 120 km. The energy required for operation of the vehicle was determined simplified calculation based on basic physical equations for calculating the energy performance of engine vehicles (Table 1).

Table 1. Energy and sampling intensity in terms of charging electric vehicles in selected categories for a single driving cycle

Category of vehicle	Stored energy of vehicle (range 120 km)	Charging current		
		t = 8 hours	t = 4 hours	t = 0,5 hour
A	13,3 kWh	5,1 A	10,2 A	81,6 A
B	30,0 kWh	10,2 A	20,4 A	163,2 A
C	41,1 kWh	14,2 A	28,4 A	227,2 A

In a similar way, was calculated for the stored energy and the expected nominal voltage of 350 V battery charging current for three different charging times. These times have been set for probable modes:

- t = 8 hours - long-term (e.g. night) charging electric vehicles by low current at full capacity battery,
- t = 4 hours - charging electric vehicles during the day (e.g. a vehicle parked during working hours around lunch breaks or business meetings),
- t = 0.5 hours - so-called fast charging as during a long journey stops.

2.2 Estimation of the amount of electricity needed to charge the vehicles in the region of Ostrava

The basic premise solutions and designing active networks of charging stations is to create conditions for adequate consumption of electrical energy in repetitive cycles, which can be linked to the cycles of the daily load diagram distribution network.

The second prerequisite for successful application of active power stations in the area is its positive effect on the network. The solution must be open enough to allow even the use of non-traditional and renewable energy sources.

Types of electric vehicles operated in the Ostrava region, in accordance with the breakdown as shown in Table. 1 could look under table. 2.

Table 2. The expected number of electric vehicles in the Ostrava region in the period 2015 - 2020

Category of vehicle	Number of vehicle (%)	Number of vehicle (pc)
A	70	17 500
B	25	6 250
C	5	1 250

On the basis of this distribution can be calculated by energy stored in vehicles for daily cruising range of 120 km, which would correspond to daily charging cycle. Stored energy in all vehicles,

$$\begin{aligned}
 W_{ak(celk)} &= N_{(A)} \cdot W_{ak(A)} + N_{(B)} \cdot W_{ak(B)} + N_{(C)} \cdot W_{ak(C)} \\
 &= 17500 \cdot 13300 + 6250 \cdot 30000 + 1250 \cdot 41100 = \\
 &= 471,6 (MWh)
 \end{aligned} \tag{1}$$

It can be assumed that the actual status of distribution network in the Ostrava region, in terms of its design and in terms of possible sources would not be enough to cover the requirements of electromobility in a certain period of time (increase output by about 42 MW). The solution shows the necessity of using active charging stations.

3 Concept of charging stations for electric cars with minimal impact on the supply network

The concept of charging stations must ensure that the energy required for charging electric cars were available regardless of network options, such as at the time of decommissioning the vehicle and its charging even during peak sampling network.

Another key requirement for active charging station is the suppression of reverse effects of different types of semiconductor chargers, which will additionally among themselves independently connected to the distribution network at different times of the day. A possible structure of such network shows Fig. 3.

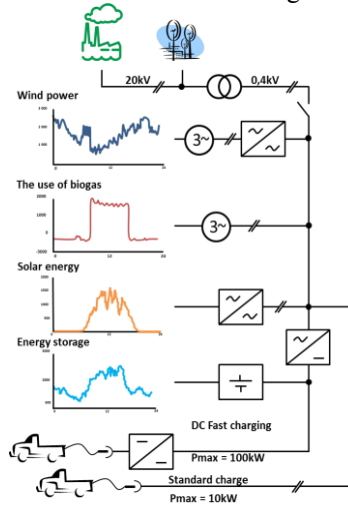


Fig. 3. The integration active charging station into the structure with the accumulation of renewable sources [4]

A distinctive feature is the use of multiple sources of energy, which significantly reduces the particular peak load distribution networks. The typically is utilization non-traditional and renewable resources in the temporal distribution of energy to the grid along with energy storage in the charging station for the network creates a stabilizing element. Involvement of active charging stations, in addition to its stabilizing function to perform a variety of additional features:

- minimizing the consumption of reactive power from the network and the conditions EMC as network traffic and RF interference,
- achieve the highest energy conversion efficiency for all types of converters,
- remote control charging modes of dispatching (e.g., discontinue charging vehicles from the network and go to the charging of the battery charging station),

3.1 The basic structure of the active charging station

For a conventional vehicle charging typically uses its own charger connected directly to the AC grid. Charging power usually does not exceed 10-15 kW. This simple method of charging usually causes problems with increased power draw in the 3. and 5. harmonic, which is due to the concept of switching power supply in the electric charger. This concept does not allow for the fast charging.

The process allows fast charging the vehicle's battery in less than 1 hour, which is a charging current of up to several times larger than the current amp-hour capacity battery. A basic comparing size of these currents is performed in the Table. 1.

Block diagram of the active charging station with AC output is shown in Figure 4.

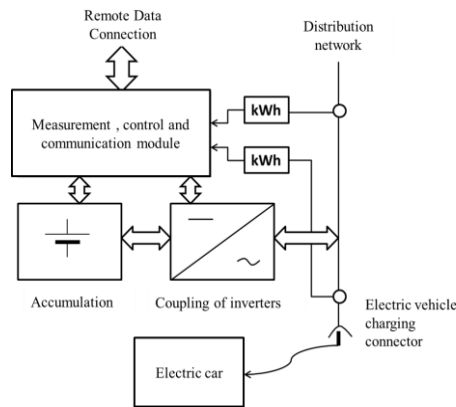


Fig. 4. The structure of the charging station for electric vehicles

The concept charging station uses the parallel connection of storage to the power supply charging socket. The basic power blocks are bidirectional inverter, which allows bidirectional power flow between the network and the battery. This arrangement allows to separate electric charge only from the network as part of the common charging a battery charging station, and part of the network and battery charging from the network in the event that the charging socket is not connected to an electric vehicle. The whole system is controlled by a unit of measurement, control and communication. The power section of active charging station must therefore satisfy the following requirements:

- high dynamic management of operating conditions,

- four-quadrant operation relative to the grid,
- active filtering reversing the effects of electric chargers,
- high efficiency of conversion of electrical energy.

This solution enables the charging station not only cover the needs of electric vehicle charging power when it is on the network by the lack of drawing power from the auxiliary battery charging station is active, but also allows for convenient control inverter suppress the negative feedback effects of electric vehicle charging on the power grid as the basic harmonic reactive power supply, and in the higher order harmonics. The advantage of the solution is also that in the case of energy shortage in the auxiliary accumulator station can only operate as a so-called parallel active filter. [5]

4 Conclusion

The paper deals with active charging stations for electric vehicles in the event of mass deployment. This type of charging stations is not yet used by us, so it's no experience with operation and impact on the power grid. The aim is to create such an active station, which is in equilibrium with the opportunities and needs daily load diagram at the site. Therefore, the whole calculation is based on the need for continuous cycling of energy in one day.

One important result of this analysis is the educated guess that the current distribution network in the Ostrava region in terms of its design and in terms of possible sources would not be enough to meet the requirements of electromobility in a certain period of time (increase output by about 42 MW). The solution shows the necessity of using active charging stations. In connection with an appropriate geographical distribution of active charging stations in both locations with their high density in cities and in areas that are likely to be filled in infrastructure passageways between the larger sites, caused by the use of active charging stations new opportunities for management and operation of distribution networks.

The work was performed with the support of Student Research Project Grant Competition SP2013/137.

References

1. <http://www.hybrid.cz/clanky/rychlonaobjeci-stanice-pro-elektromobily>
2. Barešová, H.: Effect of electric charging stations for electric grid. DP VSB-TU Ostrava. Ostrava, 2011
3. <http://www.mvcr.cz/clanek/centralni-registr-vozidel-865510.aspx?>
4. Benger, R., Heyne, R., Haubrock, A., Beck, HP: Sustainable Fast Charging Stations for Electric Vehicles. In 5 th International Renewable Energy Storage Conference (IRES 2010), Berlin, Germany
5. Korenc, V. Bubl, T.: Active filters domestic production, Proceedings of the XXVI. Conference on electrical drives, Plzeň 1999

Power quality measurement

Michal Reguľa

University of Žilina, Faculty of Electrical Engineering, Department of Power
Electrical Systems, Univerzitna 1, 01026 Žilina, Slovak Republic,
`michal.regula@fel.uniza.sk`

Abstract. This paper deals with the power quality measurement in laboratory environment on 22 kV power line model and measurement in manufacturing company focused on engineering manufacture, too. The first part of the paper presents especially problems concerning on harmonics in 22 kV power line model. The next part includes problems of measurement of qualitative indicators in the manufacturing company. The results obtained from measurements are simulated and analyzed and are assessed following the standards.

1 Introduction

In the generation, transmission and distribution of electric power are used many modern technologies and this brings new actual problems. As a result of modernization, the problems not observed in the past occur in electrical networks. Producers, distributors and consumers of electricity are becoming more concerned about the power quality. The power quality involves a large number of different types of faults in distribution systems. Power system resources and appliances are interconnected and the system elements affect each other. An incorrect operation of equipment can have serious impacts on other devices operating in the network [1], [3], [4].

The power quality is closely linked to economic impact on suppliers and consumers of electricity, as well as manufacturers of various appliances. We are often hearing about the price of non-delivered electricity or about the price of failure in operation due to reduced quality of electricity. Such failure may considerably affect economy of company. Definitely we can say that every problem connected with a variation of the voltage, current or frequency which causes a failure or incorrect operation of the equipment, it is the problem of power quality. Therefore, in this article we deal with the spread of higher harmonics through the distribution line model, their generation and effects to other customers [1], [2], [4].

Voltage or current harmonics are voltages or currents with sinusoidal waveform and a frequency that is an integer multiple of the fundamental frequency at which a network is designed to operate. There may not be only harmonics with an integer multiple of the fundamental frequency in a network but harmonics whose frequency is not an integer multiple can also occur. These are called interharmonics and can be generated by various devices, e.g. static frequency converters (inverters), cycloconverters, inverter cascades, induction motors, arc welders or arc furnaces.

The major negative effect of current harmonics is that the RMS value of a non-sinusoidal current is higher than the RMS of its fundamental component at which the equipment is designed [1], [4].

Harmful effects of harmonics can be divided as follows:

- Short-term – are associated with failure and malfunction or decreasing the quality of equipment operation caused by incorrect zero-crossing detection.
- Long-term – are essentially thermal effects. These occur after a time period of more than 10 minutes.

The most severe negative effects of harmonics in the field of energetics are:

- Improper function of control devices.
- Additional power losses in capacitors and rotating machines.
- Malfunction of telecontrol signals and other network signalization devices or protection relays.
- Thermal conditions impairment.
- Occurrence of undesirable resonances.

The most severe power quality problems are voltage sags, swells, interruption, harmonic and flickers. Failures due to such disturbances create high impact on production cost. Especially, modern industrial equipment is more sensitive to voltage sags. STN EN 50160 defines a voltage sag as a short term reduction in voltage magnitude at a value in the range 5 to 90% of the supply voltage. Even the shortest voltage sag can cause serious equipment damage, interruption of production cycles and thus financial losses as well. There are several ways how to compensate voltage sags and swells. One of these is to introduce dynamic voltage restorer, which belongs to FACTS [1], [5].

2 Analyses of power quality in a power line model

In laboratory environment conditions at the Faculty of electrical engineering of University of Žilina there is a three-phase 22 kV power line model. All the measurements of the harmonics and sags propagation were performed on this model.

Considering model construction, I discovered marked deterioration in total harmonic distortion of power voltage at higher leakage current. It was particularly the module with parameters of electric line representing the length of 10 km. It is caused by the fact that there were used coils with ferrite cores ETD 59 made of SIFFERIT material for inductance in modules. In the case of this material the core saturation occurs at the magnetic induction value of $B_{\text{sat}} = 0.49$ T. After exceeding this limit by means of the current increasing the generation and propagation of harmonics are expected because of the iron-core nonlinearity as it can be seen in Fig. 1 right-hand side.

If the model is low current loaded, as for example $\frac{1}{2} I_{\text{MAX}}$, the size of harmonics is at a lower level. Therefore I can maintain that at model loading up to $\frac{1}{2} I_{\text{MAX}}$, there is no harmonics generation and just harmonics from PCC are transmitted to the model. Harmonics spectrum can be seen in Fig. 1 on the left side. For comparison $THD_U = 2\%$ ($I = 1$ A), $THD_U = 5\%$ ($I_{\text{MAX}} = 2$ A) and $THD_U = 8\%$ ($I = 2.5$ A). As it is evident from the FFT, there occurs mainly to generating 3th harmonic.

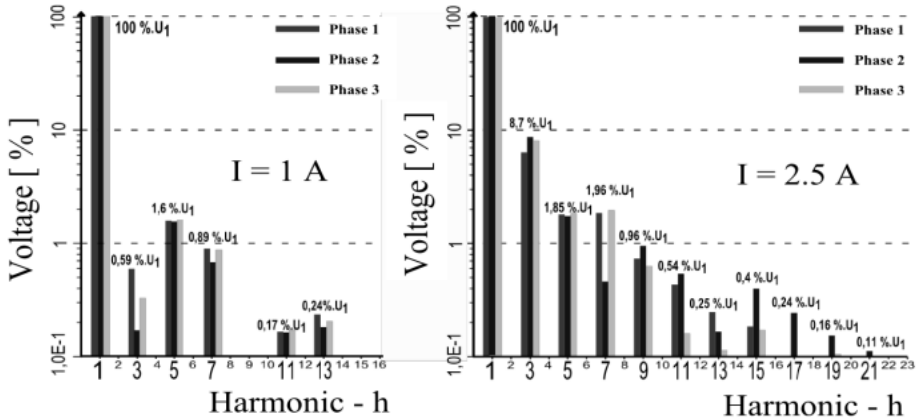


Fig. 1. FFT analysis of power line voltage for module length 10 km

I also explored influence of the interference source (rectifier) on other customers connected in distribution system. According to connection in Fig. 2, I monitored THD_U at three consumption places (Load A,B,C) and in PCC.

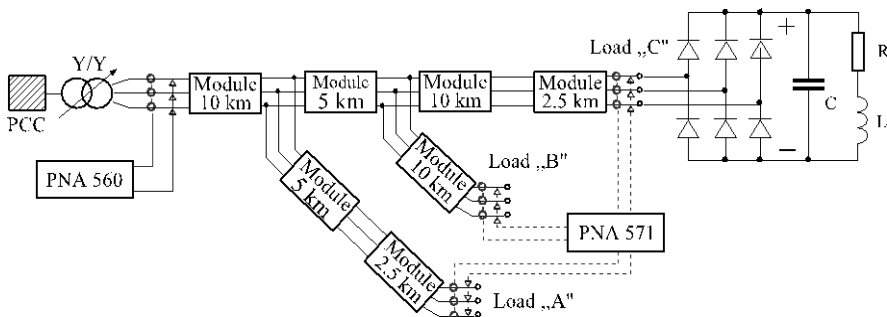


Fig. 2. Scheme of the distortion measurement

The resulting values of measurement where as an interference source was connected 6-impuls rectifier are shown in Table 1. The values show that the nearer the measurement point (customer) is to the interference source and farer from the connection point PCC, the greater influence interference has.

In PCC is distortion the smallest ($THD_U = 2\%$). This distortion was brought from superior system. In the point Load „A“ is $THD_U = 8\%$ and in Load „B“ is $THD_U = 11\%$, because Load „B“ is nearer to the interference source.

Table 1. The values of THD_U at various consumption points

Rectifier- „R“	„PCC“	Load „A“	Load „B“	Load „C“	
THD _U (%)	phase 1	2.33	8.02	11.14	16.02
	phase 2	2.08	7.28	10.91	16.19
	phase 3	2.46	7.75	10.88	16.93

3 Power quality measurement in company

Measurement in selected manufacturing company was carried out with power network analyzing instrument Dewetron DEWE 571. By means of this device there was performed one-week measurement in one of seventeen transformer station within the manufacturing company. Specifically it was transformer station marked as TS5. At this transformer station exceeding marginal qualitative indicators defined by standard STN EN 50160 was expected. In this point all administration offices are connected from one bus (economic department, projection,..) and also the biggest production hall within manufacturing company.

One of the core indicators is voltage. During monitoring period RMS voltage values did not exceed any of marginal values defined by standard as it is shown in Fig. 3.

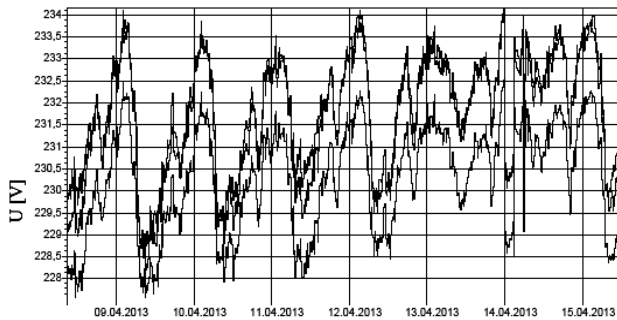


Fig. 3. Time course graph RMS voltage values (phase to ground)

Maximum measured voltage value was $U_{MAX} = 234.63$ V and minimum $U_{MIN} = 228.28$ V. Trigger (which was set for recording voltage sag on -10 % out of $U_N = 230$ V and overvoltage on +10 % out of $U_N = 230$ V) did not make any record during monitoring term. As it can be seen in Fig. 4 values THD_U and THD_I also did not exceed values defined in standards.

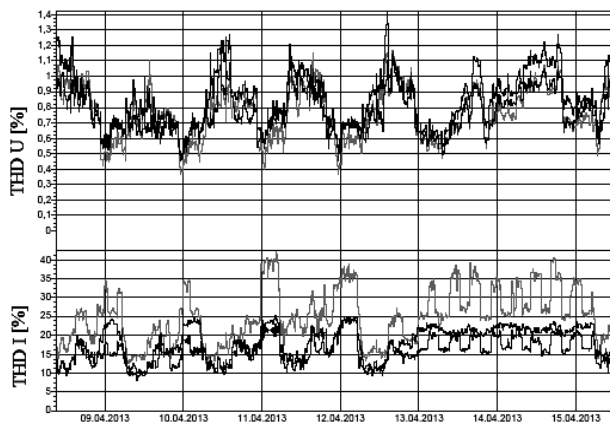


Fig. 4. Time course graph of THD voltage and current values

In the spectrum FFT (Fig. 5), there were not observed any long-term exceeds of values in accordance with standard STN EN 61000-2-4 class 2 [6]. However, in the graph we can see that short-term measured values exceed maximum limited values for 15th harmonic and what concerns 21st harmonic, the values are at the peak. The measurement results confirm that passive filters up to 13th harmonic in transformer station TS5 can eliminate harmonic effectively enough, however harmonic higher than 13th (e.g. 15, 17, 19, 21,...) are not eliminated.

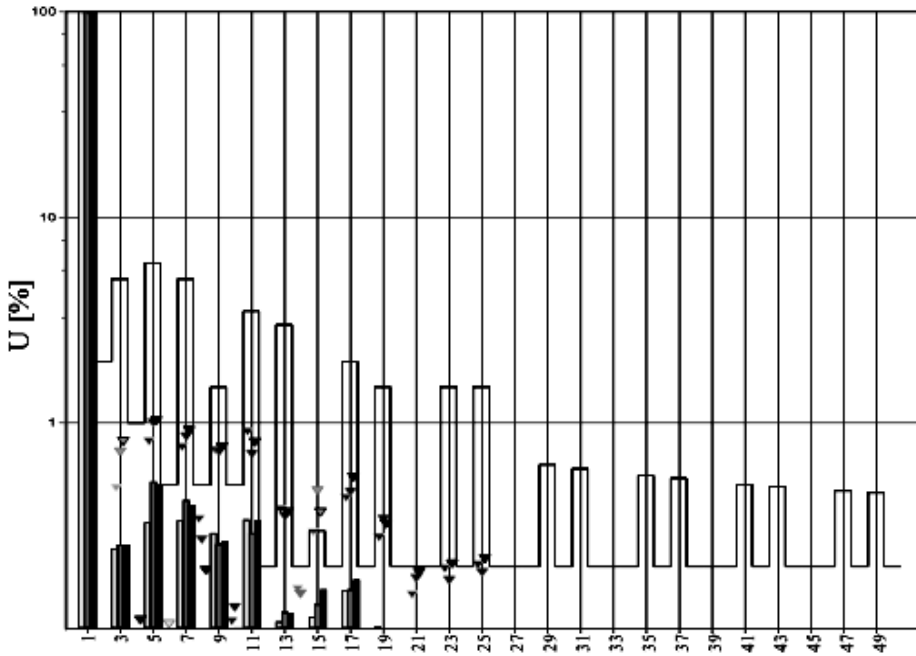


Fig. 5. FFT analysis of voltage in the company

4 Conclusion

There was conducted also other trials of harmonics and voltage sags transmission on 22 kV power line model. As distortion sources were used various types of loads, for example electric machines, rectifier loaded with different combination of RLC passive elements, compact fluorescent lamps or at earth fault,.. Obtained results were implemented into programming environment Matlab/Simulink. Output results of simulations were compared with measured results. Observed defects were in simulation removed with using conventional methods. However, in terms of paper range, it was not possible to include all measurements and simulations at 22 kV power line model.

Measurement in manufacturing company was performed on assumption of worse voltage quality in that place. This measurement did not confirm exceeding the mar-

ginal values specified in mentioned standards. From a viewpoint of another research it would be more appropriate to carry out longer term measurement (at least one month) and focus on measurement of harmonics higher than 13th. Furthermore, it would be also appropriate to monitor whether there do not occur short-term and steep voltage changes (peaks and notches). After such measurement we could deduce more reliable conclusion about power quality in the chosen company and suggest solution of the problems in case of identified defects.

Acknowledgment



**PROGRAM
CEZHRANIČNEJ
SPOLUPRÁCE**
SLOVENSKÁ REPUBLIKA
ČESKÁ REPUBLIKA



**EURÓPSKA ÚNIA
EURÓPSKY FOND
REGIONÁLNEHO ROZVOJA**
SPOLOČNE BEZ HRANÍC

Project: Cooperation between the University of Žilina and VSB-TU Ostrava at the improving quality of education and training of researchers for the field of electrical engineering, ITMS 22410320029.

References

1. Altus, J., Novák, M., Otčenášová, A., Pokorný, M.: Elektromagnetická kompatibilita elektrizačných sústav, EDIS ŽU v Žiline, Žilina, 2004.
2. Drabek, P., Kus, V.: The Education of EMC at the UWB. In: IEEE EPQU 2011 - ELECTRICAL POWER QUALITY AND UTILIZATION, Lisboa, 978-1-4673-0378-1, Pp 1-6.
3. Drabek, P., Kus, V.: EMC Issues of Low Frequency Interference of Power Electronic Converters. In: Applied Mechanics and Materials Vols. 284-287 (2013) pp 2516-2520. doi:10.4028/www.scientific.net/AMM.284-287.2516.
4. Dugan, R. C., Santoso, S., Granaghan, M. F. Mc, Electrical Power Systems Quality, 2nd ed., The McGraw-Hill Companies, Inc., NY, USA, 2003.
5. STN EN 50160 (33 0121), 2011, Charakteristiky napätia elektrickej energie dodávanej z verejnej elektrickej siete
6. STN EN 61000-2-4 (33 3432), 2003, Elektromagnetická kompatibilita (EMC), Úrovně kompatibility nízkofrekvenčních rušení šířených vedením v průmyselných podnicích

Complaints about the quality of electricity in Northern Moravia

Petr Rozehnal

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.rozehnal1@vsb.cz

Abstract. Electricity is the most widely used energy used in industry as well as in homes. For efficient use of energy is the need of good quality. When electricity supply are bad parameters, has become a situation where the complaint on power quality. Suppliers of electricity this complaint record and at the same time we have to deal with. Essential for addressing complaints on the quality of electric power is its location, and then complying with the required parameters for power quality, which are given in EN 50160th.

1 Introduction

Electricity is in the lives of the people part of everyday life. The use of electricity in industry and households, it is necessary to transport electricity from point of manufacture to point of consumption. The advantage of electric power is the ability to transport from point of manufacture to point of consumption and also on long distances, where it can be converted to another type of energy.

With the ability to transport over long distances, it is necessary to transport electricity from point of manufacture to point of consumption with minimum losses, with no power outages and also to sinusoidal supply of electrical energy was not deformed or otherwise damaged to various interference or faults in the network.

The problem with the supply of electricity to customers is the possible disorders that may be caused by faults in the network or various by interference. In case of failure or not delivered electricity required quality generated by the electricity consumer's complaint.

Distribution companies complaint to solve power quality in the shortest possible time. In northern Moravia is reported the number of complaints on the quality of electric energy distribution companies must deal with. For distribution companies, information on complaints, such as location, cause, contact the customer, the date and time of the complaint on power quality. From these data the distribution of complaint solves the power quality.

2 The reasons for of complaint to power quality

The cause of complaint electricity by electricity consumers may be several. Among the reasons that may lead to complaint for power quality, harmonics include the creation of a network, fluctuations in voltage, voltage unbalance and power interruptions to customers.

All these causes that lead to the emergence of complaints about the quality of electricity, can lead to large losses for both the electric power customer and as a supplier of electricity.

All complaint arising from the power quality must be verified energy supplier, who will determine whether there is a legitimate reason for complaint on power quality. The justification of the complaint to power quality electricity supplier informs the customer of electricity and in the event that a complaint is justified and must take corrective action. All power quality parameters, which are compared in reclaiming the power quality, are given in EN 50160. [1]

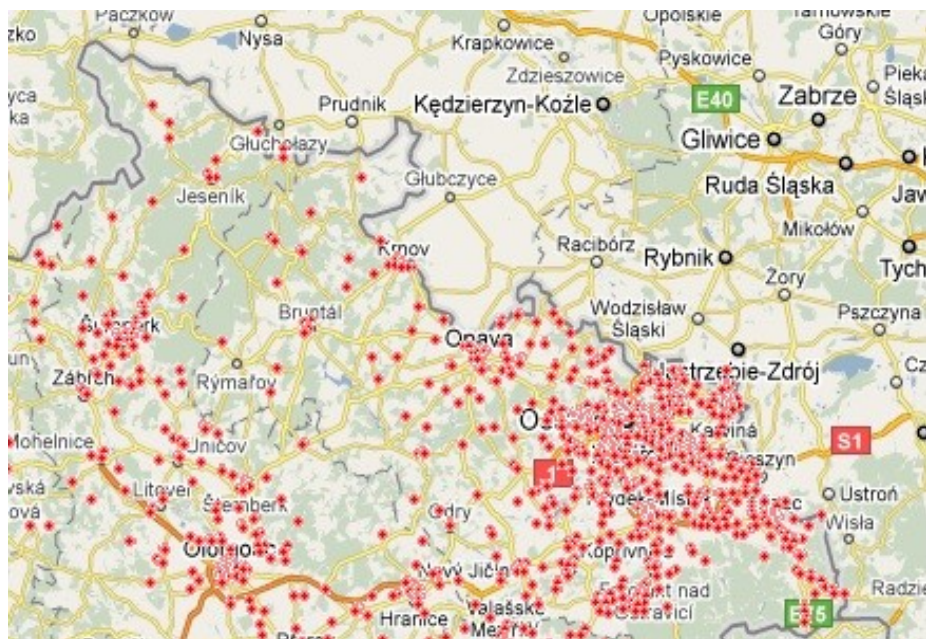


Fig. 1. Map of complaints about the quality of electricity in 2004-2011 [1]

Figure 1 shows the selected sites complaints to power quality in northern Moravia. Distribution companies all complaint to solve power quality.

2 Evolution of the number of complaints in Northern Moravia

Database of complaints about the quality of the electrical energy includes all complaints that have been reported since 2004. In northern Moravia, was, from 2004 to mid-2013 reported a total of 3546 complaints about the quality of electrical energy. The following table shows the number of complaints in the years 2004-2013.

Table 1. Complaint numbers from 2004 to mid-2013

Year	Number of complaint
2004	132
2005	67
2006	287
2007	358
2008	499
2009	477
2010	496
2011	444
2012	553
2013*	233
Total	3546

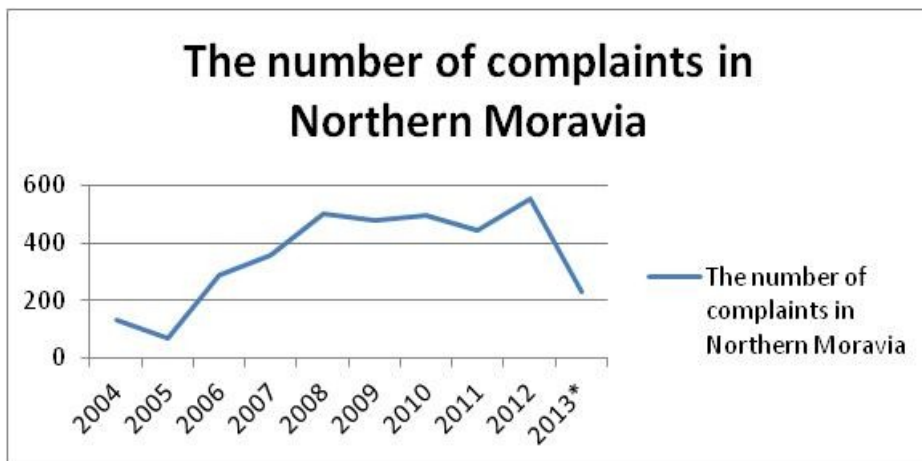
* Only the period from January to June

From Table 1 it can be seen the number of complaints in the period from 2004 to mid-2013 in North Moravia.

As you can see, the number of complaints about the quality of electricity in Northern Moravia in the individual years varies. The lowest number of complaints about the quality of electricity in 2005, when it was reported a total of 67 complaints. Against this are the most complaints on the quality of electricity in 2008, which saw a total of 499 complaints. But generally, it can be seen that the number of complaints since 2008 has not changed significantly and ranges from 450 to 500 complaints per year. From Figure 2 shows an obvious increase in the number of complaints on the quality of electricity in the period 2004-2008, when in 2004 there were 132 complaints in 2008 were a total of 499 complaints about the quality of electricity in Northern Moravia. From 2008 to 2011, but already the number complaints on the quality of the electrical energy are similar. [2]

In 2013, it was not recorded in northern Moravia total of 233 complaints about power quality. This number reflects the total number of complaints about the quality of electricity in Northern Moravia could be at the end of 2013 around 500 again

The following graph shows the development of complaints to power quality.



* Only the period from January to June

Fig. 2. Evolution of complaints about the quality of electric power from 2004 to mid-2013

Figure 2 shows the number of complaints about the quality of electrical energy. The figure shows a similar incidence of complaints about the quality of electric power from 2008 to 2012. The number of complaints about the quality of electricity in 2013 suggests that the total number complaints on power quality will be similar to past years.

3 Justification of the complaint on the quality of electricity in the Czech Republic

Justification of the complaint to power quality is very important. Distributor of electrical energy is required to deliver high-quality electricity. All parameters that are required to keep electricity distributors are listed in ČSN EN 50160th In this standard are also given different compatibility levels for all monitored parameters of power quality.

Table 2. Justification, of the complaints on the quality of electricity in the years 2004-2012.

Year of Complaint	2004	2005	2006	2007	2008	2009	2010	2011	2012
Justified complaints	0	1	146	175	155	127	46	59	0
Unjustified complaints	0	10	222	271	259	332	409	390	2276
Unspecified type complaints	462	322	974	1252	1340	1498	1741	1603	0
Number of complaints in	462	333	1342	1698	1754	1957	2196	2052	2276

Table 2 shows the number legitimate and unjustified complaints to power quality. The number of complaints about the quality of electric power is evident that the number of unjustified and justified complaints on the quality of electricity is growing each year.

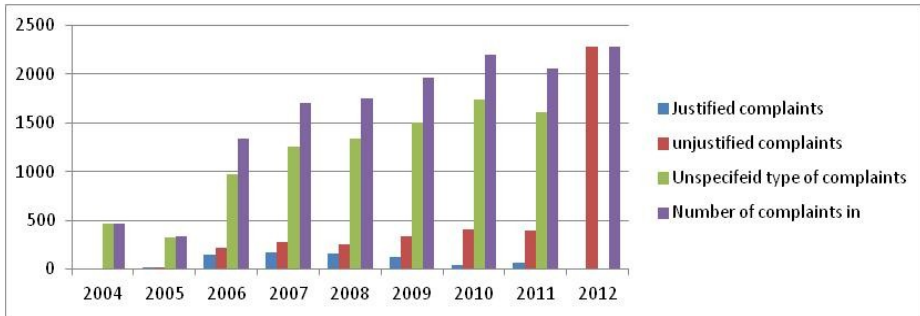


Fig. 3 Evolution of entitlement power quality

From Figure 3 shows the growth of unjustified complaints for power quality, because while in 2006 there were 222 complaints about unauthorized power quality, in 2009 there were 332, in 2010, reported 409 complaints of irregular quality of electricity in 2011 were reported 390 complaints on unauthorized power quality. [2]

Eligibility for power quality, I did not address in this article, because the database for the legal half of 2013 came at the beginning of July and I did not make it into this article evaluated.

4 Conclusion

This article summarizes research complaints on the quality of electricity in northern Moravia from 2004 to mid-2013. The reason for the occurrence of only is half of the 2013's edition of this paper.

In the first part of this post are the possible causes and location complaints on the quality of electricity in Northern Moravia. From this perspective, it is quite clear view of the location of individual complaints on the quality of electricity in Northern Moravia.

Other sections in this article are clarified number of complaints in Northern Moravia during the period from 2004 to mid-2013. In this period there were a total of 3546 complaints about the quality of electrical energy. The total annual views are evident that the number of complaints about the quality of the electrical energy increases.

In the last section of this paper summarizes the number of justified complaints on power quality. This section is limited by the end of 2012 for evaluation of claims for 2013 I missed because of the issue of this paper. From the perspective of justification of the complaint to power quality were, the number of unjustified complaint. In 2012, all complaints relating to the quality of electric energy, all complaints relating to the quality of electrical energy assessed as unjustified.

5 Acknowledgment

This work was supported by the Czech Science Foundation (No. GA ČR 102/09/1842) and by the Ministry of Education, Youth and Sports of the Czech Republic (No. SP2013/137).

References

1. Rozehnal. P, Unger J., Krejci P.: Measurement of selected quality parameters of electricity in places complaints to power quality, Conference EPE2013, Kouty nad Desnou 2013, ISBN 978-80-248-2988-3, p 639-642
2. Rozehnal. P, Unger J., Krejci P.:Complaints about power quality and reliability of electricity supply from renewable sources, Conference EMD2012, ISBN 978-609-457-260-9, Vilnius, 2012

Photovoltaic Power Plant, Extreme Thresholds of Production

Martin Smoček and Zdeněk Hradílek

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{martin.smocek, zdenek.hradilek}@vsb.cz

Abstract. Photovoltaic power plants are significant for their dependence on weather conditions. The consequences include their stochastic change to the production power that has adverse impact on electric power distribution grids. The survey focus on assess the impact of photovoltaic power plants operation on the daily load diagram with regard to output change difference of active power. Statistical methods have been employed to process a methodology in order establish extreme conditions of production power of photovoltaic power plant. Consequently these minimum and maximum extreme conditions are applied to the daily load diagram at the specific area, where the photovoltaic power plant is connected. Evaluation is based on real-time synchronous data measuring both on-site photovoltaic power plant operation and on-site electric power line output from the substation. This article describes the methodology to determine extreme conditions applicable to photovoltaic power plant production and defines its thresholds.

1 Introduction

This article deals with stochastic analysis and assessment of database comprising data obtained by measurement on a photovoltaic power plant (PVP). Specifically, it is determined extreme threshold of PVP production and subsequently effects these extreme power changes to power flowing from the substation. These changes of power flowing from the substation plant have adverse effect on supporting services provided by operators of distribution networks. Determination of extreme power changes is important for the design of devices in electrical networks too.

The evaluation is conducted by means of several individual tasks that can be split into two stages. The first stage determines methodology for set-up of extreme thresholds for production generated by the PVP and the second stage concerns analysis of impact of these extreme threshold values on the magnitude of active power flows at the place of measurement within substation plant.

The data was processed and evaluated using the Excel Microsoft office 2003 chart processor and Statgraphic XV statistical software.

2 Database of Measured Data

The data was obtained by measurement at PVP in the municipality of Starojická Lhota and on power line output (vn51) of the substation in Nový Jičín, where this PVP is connected.

The peak output of this PVP is 1.1 MWp. The data was obtained on the low voltage (NN) end. The measurement was conducted on continuous basis for one year; that was from 30.6.2010 till 29.6.2011. The measured value (observed in this survey) was represented by active power P in the course of time.

Further measurement was obtained at the substation in Nový Jičín, specifically the power line marked vn51. This measurement was conducted on the high voltage (VN) end. The line has a loop system, yet it is operated in radial system. When this measurement was conducted, the PVP Starojická Lhota was the only electric power source with significant output on the given power line. The measurement ran within various time periods from 26.7.2010 till 23.6.2011. The values measured included the active power P , once again.

Percentage evaluation and mutual comparison of data measured on PVP and at the substation point have been illustrated in Fig. 1. Columns in charts represent individual months. The last column shows an average of all months monitored.

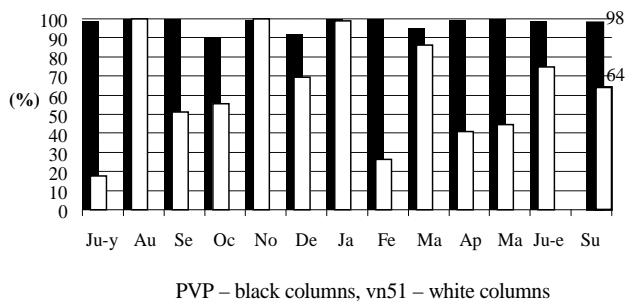


Fig. 1. Percentage comparison of data measured

More information about the database is in the source [1]. Further steps described in this article analyse data in April 2011 [3], [4].

3 Determination of Extreme Thresholds

This chapter describes the first stage of evaluated. It is define the extreme power changes of PVP production. Production of electric power by means of PVP is predisposed by the course of 24-hour daily cycle. The cycle describes a curve of Gauss distribution with maximum values in the afternoon and minimum during morning and evening hours respectively. Such distribution of electric power production is undesirable, when compared to the daily load diagram [1].

For extreme production conditions, the PVP can be assessed with respect to two extremes. These relate to the maximum and minimum production output respectively. The minimum production power is experienced during morning and evening hours, while the maximum production occurs during afternoon.

3.1 Maximum Production Extreme

Source [1] describes two entry conditions for evaluating maximum production extreme. First is such that the median of extreme daily power has to correspond with the regression curve of the maximum extreme production, another refers about correct statistical evaluation (index of determination $R^2 > 80\%$ [5]). According to these entry conditions, has been laid out methodology based on statistical method. It was conducted with division of data from the period monitored into relevant frequency categories using the so called Sturges rule. This is a rule for optimal determination of frequency categories. Subsequently, the class frequency with the maximum power produced was selected. It is show Fig. 2.

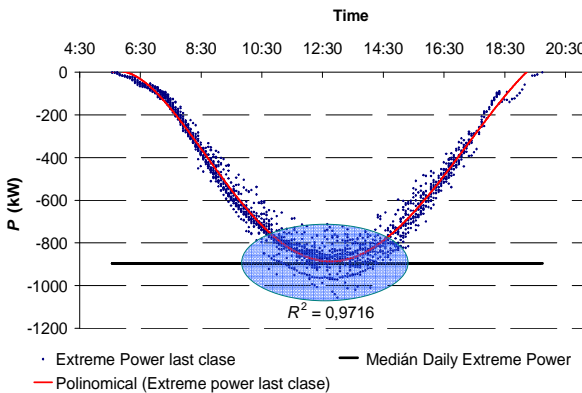


Fig. 2. Maximum frequency category of produced power by PVP

3.2 Minimum Production Extreme

Evaluation of minimum production extreme is based on similar methodology as determination of maximum production extreme. In this case, the class frequency with the minimum power produced was selected. Determine of this extreme threshold respects the diffusion radiation element that is 100% identifiable (so called guaranteed share of diffusion radiation [1]) by visual assessment in accordance with Fig. 3.

Fig. 4 shows extreme areas of PVP production. The thresholds of these production extremes are defined by two regression polynomial equations [1], which form the top and bottom envelope curves of stochastic changes in the produced power.

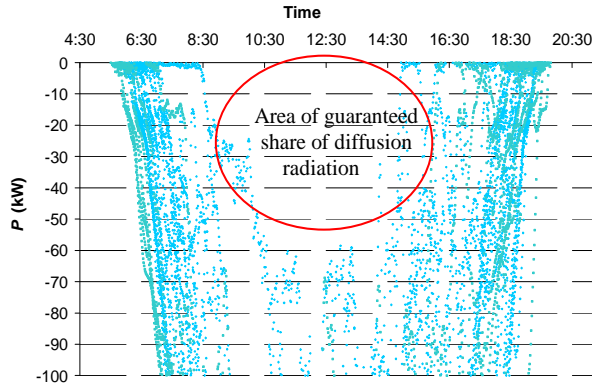


Fig. 3 Guaranteed share of diffusion radiation

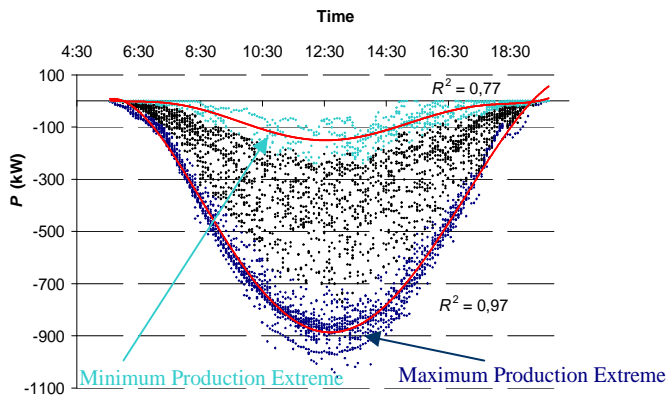


Fig. 4 Regression of both extreme production conditions

3.3 Regression analysis of the daily load diagram

Correlation fields (scatter plot in Fig. 5) of daily load diagram can be subject to regression analysis applied to enable determination of differential thresholds of changes in power produced with certain reliability. It is power flowing through the substation. Differential thresholds are determined using 95% prediction reliability levels. These prediction levels generally define the probability and range for power daily load diagram for ever individual value within specific time intervals. These thresholds are shown in Fig. 5 (purple curve – 95% prediction level, orange curve – polynomial regression of 8th degree). The regression was matched by determination index $R^2 > 86\%$, which corresponds with strongly dependent regression [5].

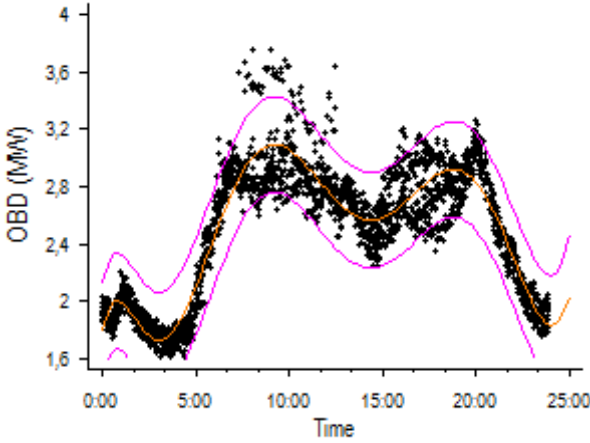


Fig. 5 Regression analyses of daily load diagram

Previous analyses have produced the difference of power generated by PVP (maximum and minimum extreme thresholds of production) and determination of thresholds for daily load diagram [2]. Extreme conditions that might occur at the substation output correspond with the sum of extreme conditions of PVP and the daily load diagram.

The Equation (1) refers to the maximum extreme power at the output from substation P_{ROZmax} . That describes the light blue curve in the Fig. 6. In this case, the daily load diagram has reached its maximum level P_{OBDmax} , whereas the contribution from PVP is at its minimum level P_{FVEmin} :

$$P_{ROZmax} = P_{OBDmax} - P_{FVEmin} \quad (MW) \quad (1)$$

The Equation (2) refers to the minimum extreme power at the output from substation P_{ROZmin} . That describes the dark blue curve in the Fig. 6. In this case, the daily load diagram has reached its minimum level P_{OBDmin} , whereas the contribution from PVP is at its maximum level P_{FVEmax} :

$$P_{ROZmin} = P_{OBDmin} - P_{FVEmax} \quad (MW) \quad (2)$$

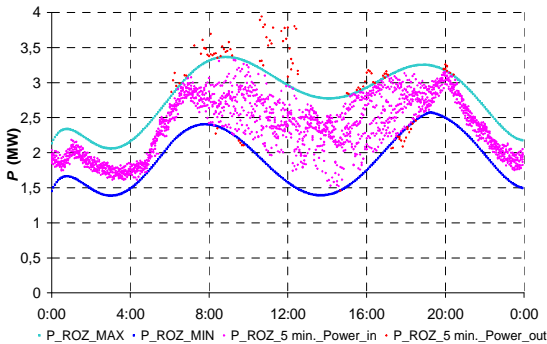


Fig. 6 Control of methodology with more than 95% reliability

In order to control of developed methodology, the values of the average five-minute power at the measurement point on the substation (pink and red scatter plot of Fig. 6) are plotted in competent thresholds of curves describing the minimum and maximum extreme power, which are based on this survey. All the data and evaluation methodology were considered with reliability level of 95 %. This particular case concerns 95.54 % of values (average five-minute output intervals) situated within the defined power thresholds. It can be therefore concluded that the threshold curves in this model respect more than 95 % of values.

4 Conclusion

This survey describes a new methodology pursuant statistical evaluation of the measured data on PVP to determine the difference of active power flowing through the substation. The objective is to observe relationship with a supporting services provided by operators of distribution networks. This study can be use to design of devices in electrical networks too. Further steps of this study consist in the application of this methodology on a full database and the potential defining discovered mutual states.

Acknowledgements

This work was supported by the Czech Science Foundation (No. GA ĀR 102/09/1842), by the Ministry of Education, Youth and Sports of the Czech Republic (No. SP2013/137) and by the project ENET (Research and Development for Innovations Operational Programme (No. CZ.1.05/2.1.00/03.0069)).

Reference

- [1] SMOĀEK, M. – HRADĀLEK, Z. Methodology for Evaluation Extreme Power Conditions of Photovoltaic Power Plants, Electrical Power Engineering 2013, VŠB-TU Ostrava, 2013, ISBN: ISBN: 978-80-248-2988- 3.
- [2] SMOĀEK, M. – HRADĀLEK, Z. “Photovoltaic Power Plants, Extreme Change of Power Difference”. The 7th International Scientific Symposium on Electrical Power Engineering, TU - Košice, 2013, in press.
- [3] SMOĀEK, M. – HRADĀLEK, Z. “Methodology for Extreme power thresholds of photovoltaic power plants”. Przegląd Elektrotechniczny 2013“, magazine of Polish electrician, 2013, ISSN 0033-2097, in press.
- [4] SMOĀEK, M. – HRADĀLEK, Z. Methodology for the statistical analysis database of photovoltaic power plant. ELNET 2012; VŠB-TU Ostrava, 2012, ISBN: 978-80-248-2926-5.
- [5] BRIŠ, R. “Statistics 1.” Study materials. VŠB-TU Ostrava, 2012, unpublished.

Monitoring of Hybrid system using LabView – Next step in development

Jindřich Stuchlý

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jindrich.stuchly.st@vsb.cz

Abstract. This paper deals with monitoring which has been developed on physical platform of family house operated in Off-grid management. Renewable sources of electric energy belong to key areas in research and development in recent years. This course leads to the design and development of autonomous power units, which are subject to specific requirements, so-called SMART GRID system or ACTIVE ENERGY MANAGEMENT (AEM), while the specific requirements are independent of external power supply, maximum efficiency of using the renewable sources, stable generation – consumption balance and opportunities of energy storage. Findings from this newly created monitoring system are presented in following chapters.

1 Introduction

Energy self-sufficient is determinates as operation of micro regions independently of the superior power system in Off-grid system. In 2011 has been established pilot project “Energy Self-sufficient Smart House Conception” to supply the family house with electrical and thermal energy using only the renewable energy sources. In this project were built up Wind Power Plant (WPP) and Photovoltaic Power Plant (PV) as an electric energy sources and solar collector as thermal energy source. This basic platform was created for researching and development in Off-grid systems at different power levels. In 2012 was developed comprehensive monitoring and remote control system, which is able to define basic operation parameters of each component such as whole system.

This software was created in LabVIEW (Laboratory Virtual Instrument Engineering Workbench), sometimes known as G-language (graphic language) that uses icons instead of lines of text to create an application. This tool of virtual instrumentation is suitable for programming measuring and analyzing signals as well as for management and visualization of technological process of complicated system. [1, 2]

2 Hybrid system description

Developed system is possible to divide into two parts, 1st part is monitoring and 2nd is remote control. The 1st part monitors basic parameters of each component in longtime period framework. The 2nd part of this system will be not described in this paper. Block scheme of Hybrid system depicts Fig. 1.

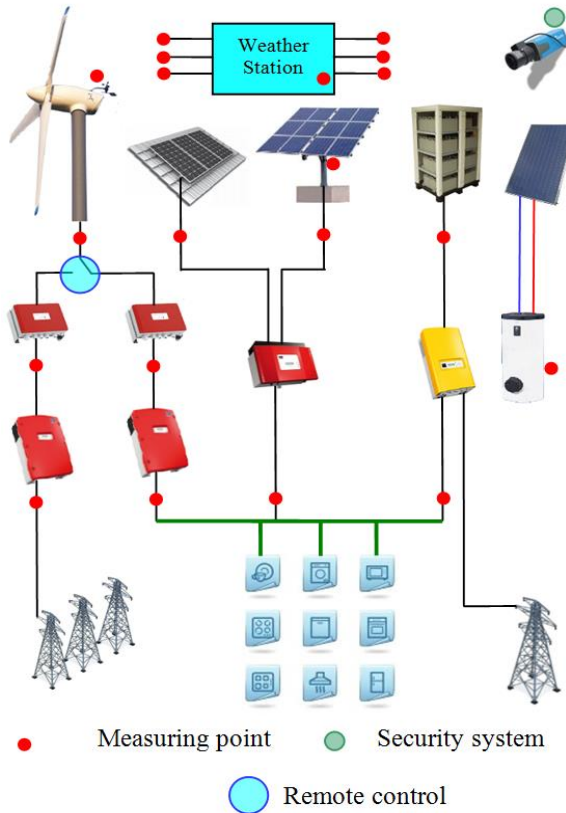


Fig. 1 Block scheme of hybrid system

In electric energy part are used voltage converters to measure voltage, current and subsequently the powers parameters. In solar energy part is used RS232 communication interface to get the basic parameters of solar collector from solar collector control unit. Another part of monitoring system is the weather station, which allows evaluating of meteorological conditions in order to compare the overall system efficiency for variety climatic. All results are saved into database for subsequently post-processing.

Fig. 2 shows the view of security system that records all sources of electric and thermal energy. Photovoltaic power plant with installed power 4 kWp is divided into two strings of 2 kWp each, the 1st is permanent installation on the roof of the building using polycrystalline solar cell technology and the 2nd is the tracker installation using monocrystalline solar cell technology. Solar Collector absorption area is 2.53 m² and

is able to safe about 1000 kW·h per year. Gondola of [wind turbine](#) is installed on the mast at 18 m above the ground and dispose with 12 kW power.



Fig. 2: Print screen from Security System with Marked Components of Hybrid System

3 Next step in development

The main goal of this project is the development of automated sophisticated dispatcher system to control the energy flows in energy self-sufficient smart house. This goal can be divided into three steps.

The 1st step is the classification of energy flows in typical household based on long-time measuring period to minimize the user's behaviour deviation and create the typical daily load curve diagram. This database will be used to set up the physics model of the typical household appliances, which will be presented by Smart Load. Smart load is RLC compilation controlled by group of contactors and relays using special digitals ports module controlled in LabView. This step will be completed by the Smart Load implementation in energy self-sufficient smart house concept.

The 2nd step represents creating the scenarios for the house (objects) behaviour that will correspond with the real situations of everyday life.

The research team will supervise and optimize energy flows between the components of energy policy with respecting the specific requirements of the Smart-grid system. The specific requirements are:

- Autonomous operation
- Using only the renewable sources of electric energy
- Equable Generation-consumption balance of electric energy
- Safety and reliability operation

The supervisor will use supporting information of the prediction model with relevant meteorological values to optimize the energy flow in the object. Predicting data is downloaded every 6 hours from the paid Meteo-server.

Each supervisor’s intervention will be monitored and recorded with detailed specification in respect of all relevant values into database using monitoring system.

During the 3rd and final phase of the project will be explored ties between the characteristic variables. This created database will be standardized and will be used for selected methods of artificial intelligence testing [3-5]. The aim of using artificial intelligence is the application for managing energy flows with respecting the specific requirements of the energy unit.

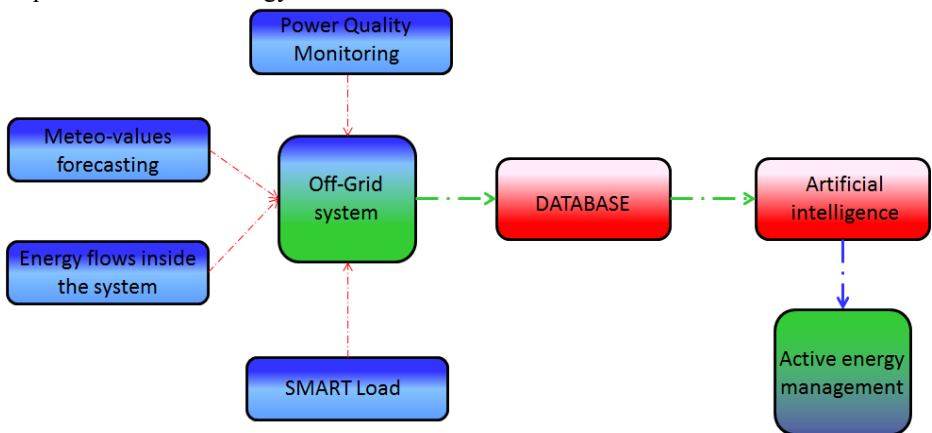


Figure 3 Flow chart of Active energy management development

The application of artificial intelligence will precede the test operation regime. Where during the test operation with applied artificial intelligence methods will be controlled the operation of self-sufficient smart house based on the normalized database. Supervisor will have permission to change the parameters via remote control management.

After the validation of artificial intelligence methods will be the last stage of 3rd step finished and the autonomous hybrid system operation can be operated by Active Energy Management. Simplified flow chart of Active Energy Management is shown on Figure 3.

4 Conclusion

Developed monitoring and remote control system has been based on the already built concept of hybrid system, which serves as a physical platform for the basic unit in the Off-grid systems. Developed software is gathering continuous measurement information, of which are evaluated individual parameters of Off-grid system. With post-processing of the measured database and long-time period measurement will be compile standardized daily load curve diagram, which will be implemented into Smart load. To use this Smart load will be the supervisor able to control generation-consumption balance as well as the character of the load, e.g. inductive or capacitive characters of the loads. After the initial testing phase of this system and collect enough information such as deliberate overload and fault states information will be implement the methods of artificial intelligence on the monitoring system database to verify the operational possibilities of this platform.

Findings from the development of this concept can be used and implemented after adjustment of input variables anywhere in the World for different voltage and power levels, either as an off-grid systems in inaccessible locations, where to building the electrical connections is unaffordable or into existing Smart-grid technologies.

Acknowledgments

This study was conducted within the framework of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 supported by Operational Program “Research and Development for Innovations” funded by Structural Funds of the European Union and state budget of the Czech Republic, project ENET – Energy Units for Utilization of non-Traditional Energy Sources CZ.1.05/2.1.00/03.0069, Students Grant Competition project reg. no. SP2013/68, and project LE13011 Creation of a PROGRES 3 Consortium Office to Support Cross-Border Cooperation.

References

- [1] LabVIEW: A Developer's Guide to Real World Integration, volume 207, July 2012, by Tylor & Francis Group, ISBN 978-1-4398-3984-2, Edited:Ian Fairweather,Anne Brumfield
- [2] Labview: For Electric Circuits, Machines, Drives, and Laboratories, volume 382, August 2002, by Prentice Hall PTR, ISBN 0-13-061886-1, Author: Nesimi Ertugrul
- [3] Cost/worth assessment of reliability improvement in distribution networks by means of artificial intelligence, International Journal of ElectricalPower & Energy Systems, Volume 32, Issue 5, June 2010, Pages 530–538, Author: Aggelos S. Bouhouras, Dimitris P. Labridis, Anastasios G. Bakirtzis
- [4] Applications of artificial intelligence in power systems, Electric Power Systems Research, Volume 41, Issue 2, May 1997, Pages 117–131, Author: S. Madan, K.E. Bollinger
- [5] Probabilistic neural network, Neural Networks, Volume 3, Issue 1, 1990, Pages 109–118, Author: Donald F. Specht, Lockheed Missiles & Space Company, Inc. Received 5 August 1988; revised and accepted 14 June 1989

Modelling of a DVR model based on DQ transformation control technique in MATLAB/Simulink

Dominik Szabó

University of Žilina, Department of Power Electrical Systems, Univerzitná 1, 01026
Žilina, Slovak Republic
dominik.szabo@fel.uniza.sk

Abstract. Nowadays, the question of the voltage quality is rapidly increasing. New technologies are introduced and we are facing many new power quality requirements. Flexible alternating current transmission systems (FACTS) are modern devices in power transmission and grid stability. The paper deals with modelling a DVR simulation model and testing a voltage sags detection algorithm based on DQ transformation. The proposed algorithm is designed to operate correctly even during some disturbances and fault conditions. Several simulations have been performed in Matlab/Simulink in order to test the function of the algorithm and the whole DVR system.

Key words: FACTS, DVR, DQ transformation, power quality, voltage sags

1 Introduction

In order to evaluate the level of the power quality, the STN EN 50160 standard was introduced, which stipulates the limits for voltage quality [2]. The most severe power quality problems are voltage sags, swells, interruption, harmonic and flickers. Failures due to such disturbances create high impact on production cost [3]. Thus, the companies are often forced to save its facilities on their own. One of the options for power quality and system stability improvement is to introduce FACTS devices.

FACTS (*Flexible Alternating Current Transmission systems*) are alternating current transmission systems incorporating power electronic-based and other static controllers to enhance controllability and increase power transfer capability [5]. FACTS controllers are able to affect and control one or several key parameters in power transmission, such as current, voltage, active, reactive power, frequency or phase. The main disadvantage of using FACTS is high cost of these devices. According to [4] FACTS can be divided into four basic types – series connected, shunt connected, combined series-series and combined series-shunt controllers.

2 Dynamic voltage restorer

Dynamic Voltage Restorer (DVR) belongs to series connected FACTS controllers. The primary function of a DVR is to mitigate voltage sags and swells but it can also perform the tasks such as harmonics compensation, reduction of transients in voltage and fault current limitation [6]. DVR is usually installed between source and sensitive load consumer which is protected. Even the shortest voltage sag can cause serious equipment damage, interruption of production cycles and thus financial losses.

In general a DVR consists of three parts (Fig. 1):

- measuring unit,
- control,
- power circuit.

The measuring unit provides voltage and current measurements. The outputs are voltage and current analogue signals (u , i), which enter the control unit. The control unit converts these signals to their digital representation using A/D converter so that they can be processed by a microprocessor (DSP – *digital signal microprocessor*). Next part is a voltage sags detection algorithm followed by compensating voltage calculation u_{com} , which is the voltage needed to be injected into the system in order to remain the load side voltage of purely sinusoidal waveform. Power section consists of a voltage source converter (VSC) equipped with a DC energy storage device and an injection transformer (TR) – *booster*.

The basic principle of DVR function is to inject or draw the compensating voltage u_{inj} to or from the supply voltage u_S in order to mitigate voltage sags or swells on the load side u_L [7]. At every moment the control algorithm compares desired voltage and actual measured voltage. The difference between these two signals is considered as a compensating voltage signal (control signal) u_{com} , which is directly proportional to compensating injected voltage u_{inj} (power circuit). u_{com} is input signal for a pulse width modulation (PWM) to control the voltage source converter. The VSC converts DC energy stored in an energy storage device (such as batteries or supercapacitors) to injecting AC voltage that is to be superimposed to the source voltage. DVR power output depends on the amount of energy that can be stored in the energy storage device. DVRs are mostly installed to protect large electrical energy consumers with the sensitive technologies and devices (2 MVA and up) connected at distribution voltage [4].

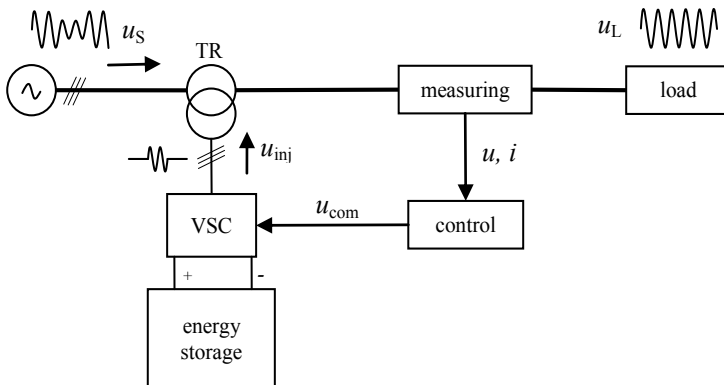


Fig. 1. Principal scheme of a DVR

3 Voltage sags detection technique based on DQ transformation

The essential requirement for DVR is to be able to operate in real-time. It means, that the whole compensating process is carried out immediately after a failure occurs without any delay. A very important factor that influences the DVR speed the most is

the reaction time of the implemented voltage sag detection algorithm. The most sophisticated DVR systems are able to react within 1 ms [7].

In this paper abc to dq0 transformation based simple control method is used to compensate voltage sags and harmonics. DQ transformation (dq0 – *direct-quadrature-zero*) is a mathematical transformation used to simplify the analysis of three-phase circuits. DQ transformation is calculated as follows [1] [2]:

$$\begin{bmatrix} u_d \\ u_q \\ u_0 \end{bmatrix} = \frac{2}{3} \begin{bmatrix} \sin(\omega t) & \sin\left(\omega t - \frac{2\pi}{3}\right) & \sin\left(\omega t + \frac{2\pi}{3}\right) \\ \cos(\omega t) & \cos\left(\omega t - \frac{2\pi}{3}\right) & \cos\left(\omega t + \frac{2\pi}{3}\right) \\ 1/2 & 1/2 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} u_a \\ u_b \\ u_c \end{bmatrix} \tag{1}$$

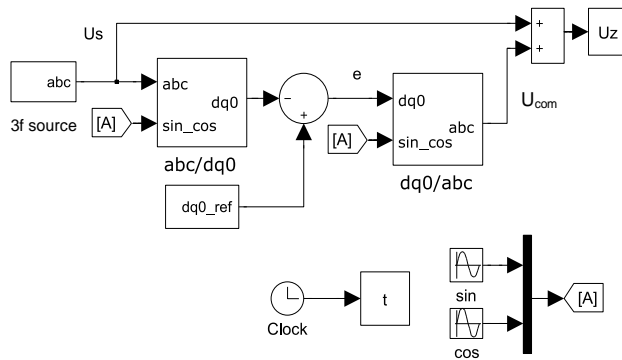


Fig. 2. Voltage sags detection algorithm using DQ transformation in Simulink

In the case of a symmetric three-phase system, the application of the dq0 transformation reduces the three AC quantities (pu) to two DC quantities ($d=1, q=0$). For unbalanced and asymmetric three phase system there is $d \neq 0, q \neq 0, 0 \neq 0$. Any deviations from the steady state condition in abc system reflect in changes of dq0 values in real-time. According to this assumption it is possible to obtain the difference between desired and instant values and thus the output compensating voltage can be controlled using PID regulators [2]. The resultant signal is converted back to abc values. The simple algorithm simulation model is shown in Fig. 2.

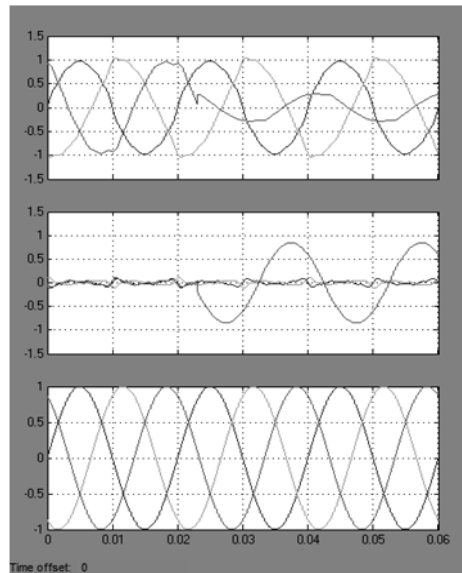


Fig. 3. Single-phase voltage sag compensating in the case of the distorted supply voltage

Fig. 3 shows three-phase distorted voltage waveform ($THD = 6.3\%$) – top picture, computed compensating voltage – middle picture and load voltage after compensation – bottom picture. In this case DVR not only compensates voltage sag in phase C but also operates as an active filter and eliminates harmonics [3].

The proposed simulation model represents only the control algorithm, not the entire DVR device. The complete DVR model is described in the following chapter.

4 DVR model in Matlab/Simulink

To prove the correct function of the algorithm it is necessary to create a complete DVR model and perform some tests. Fig. 4 depicts the model of a DVR in Simulink. Again, it consists of three main parts: measuring unit, control and power circuit.

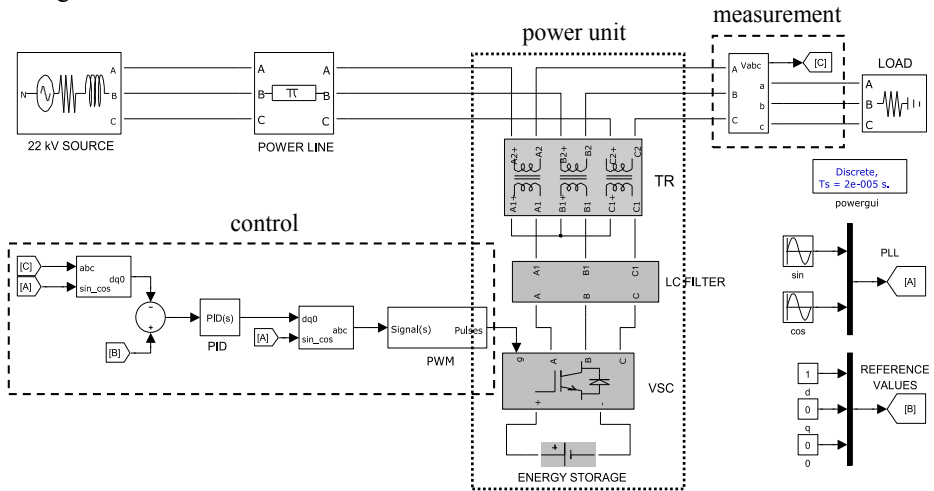


Fig. 4. DVR model in Simulink

DVR is connected at 22 kV level and between a source and a protected load. The control mechanism is described in the previous chapter. For the energy storage an ideal DC source is used. A LC filter is applied on the VSC (*voltage source converter*) output terminals to smooth the compensating voltage which is injected and superimposed to the supply voltage using series-connected three-phase transformer [4]. If a voltage sag or swell or any voltage disturbance occurs on the load input terminals, DVR reacts very quickly and computes a compensating voltage. Subsequently, this required voltage is generated by a VSC and energy storage, smoothed by a filter and added to supply voltage through a transformer. The whole process of compensation lasts less than 2 ms. The protected consumer does not perceive any of the disturbances in the electric network. DVR ensures that the supply voltage remains purely sinusoidal at any time regardless of the type of the fault.

Fig. 5. illustrates a waveform of a supply voltage during a two-phase sag in distributing network (upper picture). The sag started at 0.075 s and ended in 0.26 s. The voltage is displayed in pu values. The bottom picture shows the load side voltage

and it is clear that no modifications or just little ones have occurred. The three-phase load voltage waveform stays unaffected.

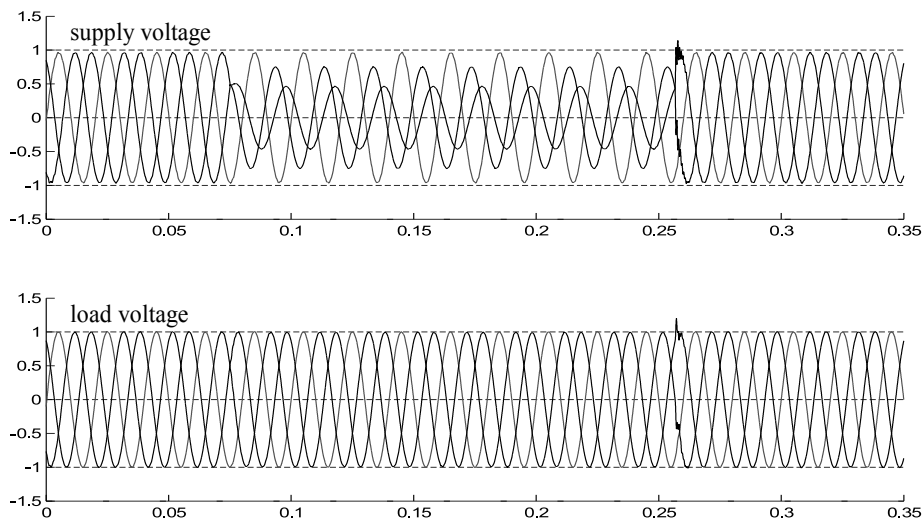


Fig. 5. Two-phase sag compensation

The maximum compensating time of a DVR (the maximum time during which a DVR is able to compensate the voltage at maximum magnitude) depends on the energy storage and its characteristics. The more energy is stored in the bank the longer a DVR will compensate. In the case of our simulation, the energy storage is ideal and inexhaustible but in a real situation, supercapacitors or batteries are applied. It means there are some limits that should be taken into a count in further analyses. Generally, the simulation tests showed the proper functionality and desired results as a whole and the proposed control algorithm is suitable to be employed to control a complex DVR system.

5 Conclusion

To sum up, the presented DVR simulation model is based on a DQ transformation control algorithm. The main advantages of this technique are its high speed, simplicity and possibility to use it even during distorted voltage waveform. On the other hand the disadvantages are that DQ transformation can be only applied in the case of three phase system and a reference sinusoidal signal is required (PLL) synchronized with the voltage before the fault. The complete DVR model was created and its correct function was verified by simulation in Matlab/Simulink. The obtained results show that, the proposed algorithm mitigates voltage sags and harmonics very fast and satisfactory.

Beside the speed of detection algorithm there are many other features that affect the DVR operation. They are, for example, the signal sampling frequency, the speed of the PID regulators, the energy storage dynamic performance, the converter

switching speed and the type of the disturbance. The correct function verification is the first step in the developing and realization of a dynamic voltage restorer.

Acknowledgements



PROGRAM
CEZHRANIČNEJ
SPOLUPRÁCE
SLOVENSKÁ REPUBLIKA
ČESKÁ REPUBLIKA



EURÓPSKA ÚNIA
EURÓPSKY FOND
REGIONÁLNEHO ROZVOJA
SPOLOČNE BEZ HRANÍČ

Project: Cooperation between the University of Žilina and VSB-TU Ostrava at the improving quality of education and training of researchers for the field of electrical engineering, ITMS 22410320029.

References

- [1] GONZALEZ, M.: *DQ transformation development for single-phase systems to compensate harmonic distortion and reactive power*. Power Electronics Congress, 2004. CIEP 2004. 9th IEEE International. p. 177-182
- [2] HEČKO, P.: *Kompenzace krátkodobých poklesů a přerušení napětí v distribuční soustavě*. University of Žilina, faculty of electrical engineering, 2010. Dissertation thesis.
- [3] HEČKO, P., OTČENÁŠOVÁ, A.: *Praktické měření kvality elektrické energie a možnost zlepšení její kvality*. International Scientific Conference Electric Power Engineering 2010, Brno, 2010, 05, 4.-6., AFC, p.: 2
- [4] HINGORANI, G. N., GYUGYI, L.: *Understanding FACTS. Concepts and Technology of Flexible AC Transmission Systems*. New York: IEEE Press, 2000. 432 p. ISBN 0-7803-3455-8
- [5] IEEE: *Proposed terms and definitions for flexible AC transmission system (FACTS)*, [online]. IEEE Xplore digital library, October 1997. ISSN0885-8977
- [6] KANTARIA, R.A.: *A novel technique for mitigation of voltage sag/swell by Dynamic Voltage Restorer*. Electro/Information Technology (EIT), 2010 IEEE International Conference. Conference publication. p.: 1-4
- [7] TANDJAQUI, M. N. a kol.: *Sensitive Loads Voltage Improvement Using Dynamic Voltage Restorer*. International Conference on Electrical Engineering and Informatics, 2011. Conference publication. IEEE Xplore digital library

Design, realization and analysis-methods for measurement heat pump energy balance

Jan Šrámek

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.sramek@vsb.cz

Abstract. This article presents a procedure for designing a measurement methodology to address to a currently very live and frequently discussed topic. That is for sure the utilisation of heat pumps of air/water type and their effective implementation in practical installations. The first chapters contain a description of the object measure and the method for designing the convenient measuring system. The measurement concerns overall energy balance of heat pump incl. heat capacity and secondary parameters associated with operation of this source by direct or indirect means. The second part of this article presents the first partial results based on initial and implemented analysis of the data measured.

Key words heating, measurement, energy balance, heat pump, coefficient of performance

1 Measure object-heat pump

The energy balance measurement was conducted within an object with intermittent utilisation. The primary heating system in this object comprises air/water heat pump only. The object is situated in the area of Beskydy, Ostravice-Staré Hamry (CZE); and its built-up area amounts to the total of 131 m². The total heat loss incurred by the object is equal to 4.5 kW.. The measurement period lasts from autumn 2012 to spring 2013.

The primary heating source comprises the air/water heat pump with the nominal output of 8 kW under A2/W35 conditions. The bivalent source is formed by an electric cartridge heater. Appliance connected within the system include a hot water reservoir and heating pipes that form the under floor heating unit.



Fig 1. Installed heat pump and heated object. (Šrámek,2012)

2 Measurement objective

The measurement focuses on two essential objectives that can be described as monitoring of full heating season and shorter periods. That will enable us retrieve and measure data for assessment of the full heating season as a whole with simultaneous monitoring of rapid changes in the heat pump on/off mode.

2.1 First and second stage

It focuses on the full heating season and determination of the average performance factor and the total consumption of heat pump, which will then enable us convert such data into e.g. actual operating costs for the full heating season.

The second stage focuses on monitoring of shorter periods and reaction of the heat pump to rapid changes in heat requirements. These conditions will be also simulated to obtain the data necessary. The measured data helps us observe the behaviour of entire system subject to measurement and the data obtained at this level can be used to find mathematical dependencies between the heat source, heat capacity and appliance (heating system).

2.2 Relation between

Standards for testing of heat pumps (ČSN EN 14 511 series) have been adopted from the European Commission for Standardisation and provide manufacturers with terms and conditions determined for testing of heat pumps. However these terms and conditions are mostly different from the actual operation of a heating system. The matters concerned are especially lower temperature gradients (input and output media temperature) during heat pump testing. [5] A typical example is the air/water heat pump with testing conditions set to 35°C temperature of output fluid. However, the system is

designed for operation at higher temperature of 40-45°C. The same case applies to the object measured.

3 Measurement engineering design and installation

The most serious problem was the installation of hydraulic components into the already finished pipeline from heat pump. Provisions for monitoring of the data necessary were based on the existing measuring equipment and components:

- Heat gauge SIEMENS Megatron 2 (metrologically proven gauge)
- Distribution networks monitor MDS-U
- Measuring bridge SIEMENS ACS tools for heat pumps
- Notebook (XP operation system)
- Data cables and converter
- MDS-U software
- ACS tool software
- MS Excel

The whole measurement system has been designed as automated. In spite of that, we had to spend the first weeks with more frequent monitoring and fine tuning of the system to eradicate some errors, especially in settings of the software equipment. Continuous download of data will enable us conduct checks during measurement and data analysis. On the basis of obtained values we can design steps for usability of actual heating system e.g. by better accumulation of the heat.

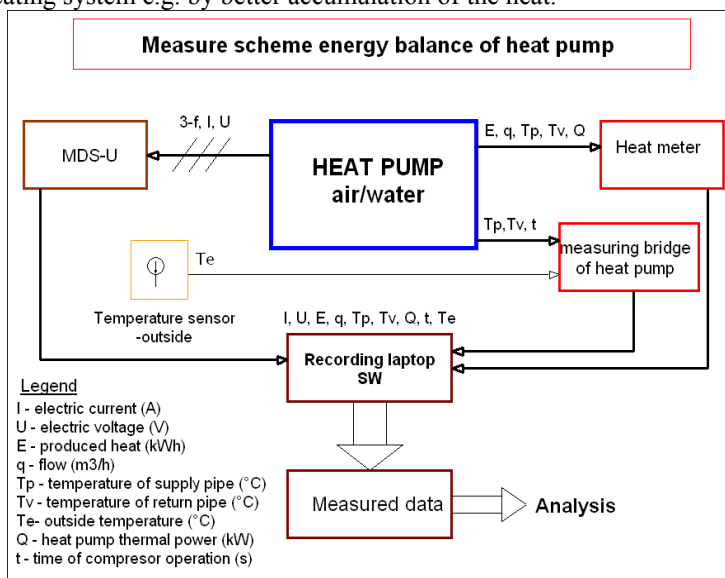


Fig 2. Measuring block diagram

On the basis of obtained values we can design steps for usability of actual heating system e.g. by better accumulation of the heat.

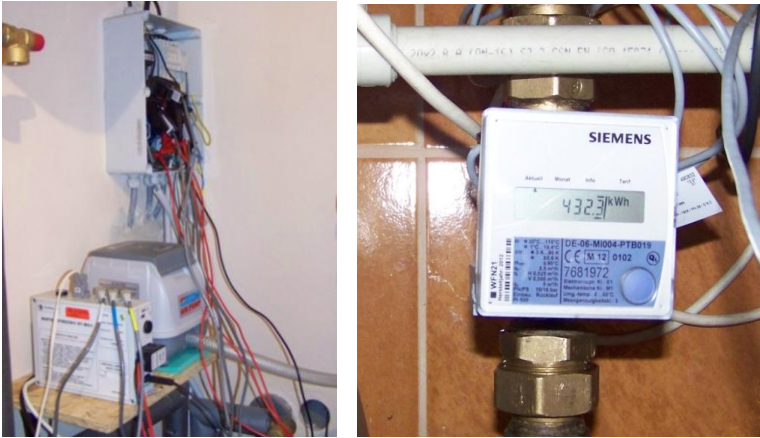


Fig 3. Installed components (Šrámek, 2012)

Practical installation of measuring components was performed at the break of September/October 2012. The process dealt mainly with integration of heat gauge into the heat pump return pipe.

4 Evaluation and first result

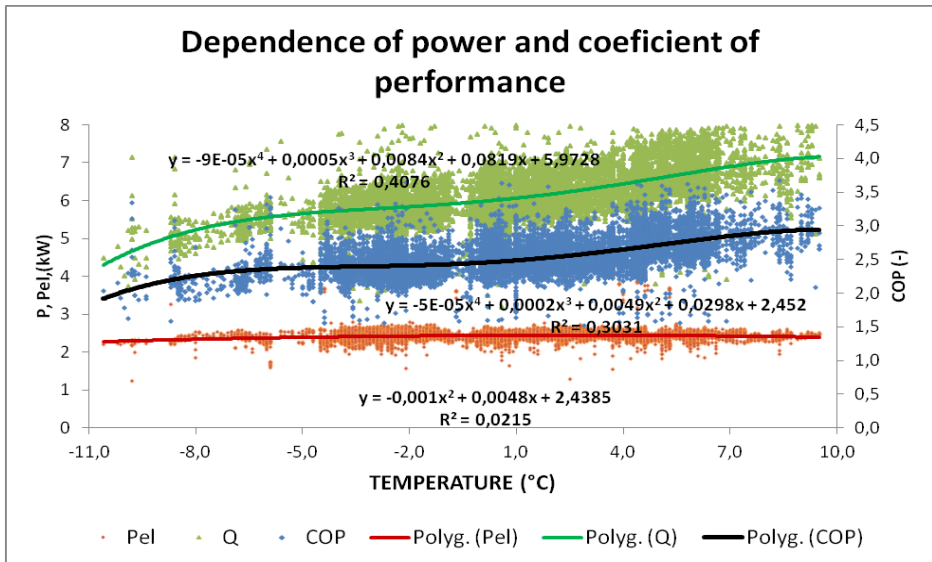


Fig 4. The dependence coefficient of performance (COP) and power according to outside temperature

The data measured is downloaded continuously during the entire process. The data analysis must be preceded by "cleaning" of incomplete data by using statistical meth-

ods. Effective processing of data, both measured and calculated, requires application of statistical methods to help us filter out extreme deflections of values incurred during transient effects.

5 Conclusion

The methodology of measurement presented in this article and the data analysis outlined here represent actual presentable opportunity to obtain data about operation of air/water heat pumps to be further developed to establish the heat asynchrony in heating of the object concerned. The objective is to obtain the data for full heating season to establish the functional dependency of heat pump output within basically the whole temperature range as required. The functional dependency, together with mode/s of heating in the object concerned will enable us innovate sizing of heat pumps to ensure more effective utilization of their options. This is also associated with monitoring of parameters of charging and discharging of the heat capacity integrated within the system. We can further work on improved effectiveness of design with respect to tank size.

The extensive data file enables derivation of various dependencies and variables in the heat pump operation, yet the process requires appropriate alteration of data file, as indicated in this articles. The output data observed in December, as shown in Fig.6, provide a truly detailed description of operation of the air/water heat pump. Subsequent derivation of functional dependency of the heat pump output will help us assess the correct selection of heat pumps with different output characteristics (stronger or weaker) with regard to the defined entry terms and conditions defined. The outlined analysis will be further finalized pursuant to complete data from full heating season. The next step will mainly concern a more detailed clearing and analysis of the data file and reduction of unreliability of dependencies illustrated here.

The outcome from this analysis will comprise especially optimization of heat pump behaviour with respect to operation and reliability aspects, reduction of investments as well as direct operating costs of heating. The proposed methodology of measurement is then aimed towards obtaining relevant data that is even measured in duplicate under some circumstances. That will provide us with the necessary data wrap.

This method has enabled us clear the data file of error and incomplete data values. Further reduction of the data file must be conducted very carefully to avoid removal of vital yet seldom operating conditions. These conditions will be also identified below.

Acknowledgements

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (No. SP2013/137) and Czech science foundation (GAČR No.102/09/1842) and by the project ENET (Research and Development for Innovations Operational Programme (CZ.1.05/2.1.00/03.0069).

Publications

1. ŠRÁMEK, Jan. HRADÍLEK, Zdeněk. MCA methods evaluation of alternative heating system application. In. Proceedings of the 13th International Scientific Conference EPE 2012 May 23-25, 2012 Brno, Czech Republic. ISBN: 978-80-214-4514-7
2. ŠRÁMEK, Jan. Accumulation of Heat Energy from Heat Pump. In. Proceedings of the 13th International Scientific Conference WOFEX 2012, Ostrava, Czech Republic. ISBN: 978-80-248-2393-5
3. ŠRÁMEK, Jan. HRADÍLEK, Zdeněk. Heat Pump En-ergy Balance Measurement. In. Proceedings of the 9th Workshop ELNET 2012, 2012 Ostrava, Czech Republic. ISBN: 978-80-248-2926-5
4. ŠRÁMEK, Jan. HRADÍLEK, Zdeněk. Methods for Measuring Heat Pump Energy Balance. In. Proceedings of the 14th International Scientific Conference EPE 2013 May 28-30, 2013 Dlouhé Stráně, Czech Republic. ISBN:

References

- [1] ŠRÁMEK, Jan. HRADÍLEK, Zdeněk. MCA methods evaluation of alternative heating system application. In. Proceedings of the 13th International Scientific Conference EPE 2012 May 23-25, 2012 Brno, Czech Republic. ISBN: 978-80-214-4514-7
- [2] ŠRÁMEK, Jan. HRADÍLEK, Zdeněk. Heat Pump En-ergy Balance Measurement. In. Proceedings of the 9th Workshop ELNET 2012, 2012 Ostrava, Czech Republic. ISBN: 978-80-248-2926-5
- [3] ŠRÁMEK, Jan. HRADÍLEK, Zdeněk. Methods for Measuring Heat Pump Energy Balance. In. Proceedings of the 14th International Scientific Conference EPE 2013 May 28-30, 2013 Dlouhé Stráně, Czech Republic. ISBN:
- [4] HRADÍLEK, Zdeněk. Energetika průmyslových a distribučních sítí. 1. vyd. 2008. ISBN 987-80-7225-291-6.
- [5] ČSN EN 14511-2 Air conditioners, liquid chilling packages and heat pumps with electrically driven com-pressors for space heating and cooling – Part 2: The conditions. 2012,24p.
- [6] BRIŠ, Radim. Statistic 1, Study materials, VŠB-TU Ostrava. 2004

Measures to achieve energy savings of lighting systems using daylight

Jan Šumpich, Zbyněk Carbol, and Karel Sokanský

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jan.sumpich, zbynek.carbol, karel.sokansky}@vsb.cz

Abstract. This paper deals with a problem how to calculate possible savings by using daylight and artificial light together. It solves calculations of artificial lighting systems dimming in the buildings which depend on the level of daylight on the working area. The problem was being discussed because the clients of lighting engineers wondered if the dimming is worth installing. The software could be used so that it meets the requirements of recommended decreasing in consumption of lighting systems in the current buildings.

1 Introduction

The effort of economic savings forces us to reduce power consumption. This phenomenon does not avoid the indoor lighting systems. In the lighting of indoor working areas is increasingly used a combination of daylight and artificial light. At present the artificial lighting systems even up to illuminance only the difference that the daylight fails to illuminate. It is done according to data from illuminance sensors. The current technical level of lighting systems allows dimming of luminaires to a constant illuminance level without significant increasing of investment costs. In these lighting systems it can be saved considerable power consumption due to daylight using.

The modeling the potential savings must be based on the level of external illumination. However, this is significantly variable. For the current calculations of daylight there is used a model of uniformly overcast sky with the intensity of illumination of 5 klx respectively 20 klx only. We designed a dynamic model based on uniformly overcast sky with variable illuminance during the day. This model can be used for the whole year for calculation of illuminance in interiors caused by daylight, without necessity to know the windows orientation to the cardinal points.

2 The procedure for predicting the energy savings

The modeling the potential savings must be based on the level of external illumination. However, this is significantly variable. For the current calculations of daylight there is used a model of uniformly overcast sky with the intensity of illumination of 5 klx respectively 20 klx only. We designed a dynamic model based on uniformly overcast sky with variable illuminance during the day. This model can be used for the whole year for calculation of illuminance in interiors caused by daylight, without necessity to know the windows orientation to the cardinal points.

This part of the paper illustrates the dynamic modeling of daylight as the base for calculating of the dimming level in artificial lighting systems and subsequent quantification of energy saving potential. In dynamic modeling there are series of aspects, which need to be taken into account. The base for dynamic modeling daylight is uniformly overcast sky as mentioned above. Due to this the influence of windows positions towards the cardinal points can be eliminated. In the calculation we do not consider only the uniformly overcast sky, but also the influence of the Sun declination changing throughout the days and year.(see Fig. 1.)

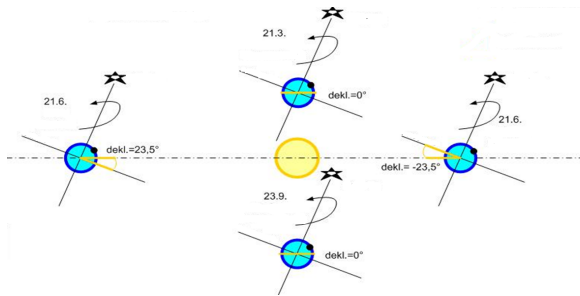


Fig. 1. Sun declination during the year

To demonstrate the calculation of daylight factors we chose a model room that is located on the 3rd floor (for the calculation itself surrounding buildings and trees are taken into account). The dimensions of the model room are 4300 x 5000 mm and height of 2670 mm. There are windows on the shorter wall. The first triple-window proportions are 1400 x 1800 mm at a height of 1030 mm above the floor and the distance from the corner of the wall is 480 mm. Second double-window proportions are 1400 x 1140 mm at a height of 1030 mm above the floor and the distance from the corner of the wall is 180 mm. The distance between the windows is 700 mm. During the calculations carried out in the model room it is necessary to determine the daylight factor, which is located on the picture (see Fig. 2.). For this calculation there was used WDL software.

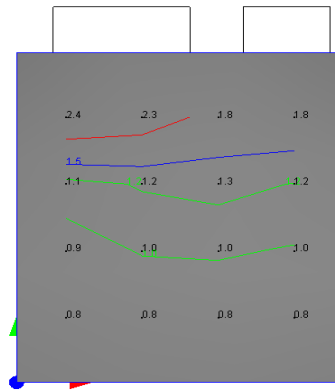


Fig. 2. The calculation of the daylight factor in the model room

For the further steps in the calculation of energy savings it is necessary to design an artificial lighting system for the model room. It is a common office (model room - see Fig. 3.) The norm (EN 12464-1) requires in this office the maintain illuminance 500 lx. As in that model room there is not specified place for visual task it was calculated so that the requirement of 500 lx is met at each computational point. The design of artificial lighting system has been done in the program WILS according to the exact dimensions of the space. In the space there are placed four surface mounted luminaires for linear fluorescent lamps (THORN - QUATTROC BODY 2x36W TC-L HFI WL6 L840).

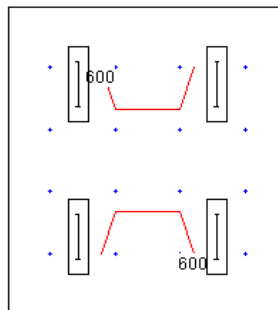


Fig. 3. The design of artificial lighting in the model room

Based on knowledge of daylight factors and artificial illuminance in the room there can be calculated a combination of the daylight and the artificial light. An example of calculating is a combination of the daylight and the artificial light for one particular time of the year. After basic calculation of day and artificial light, it is possible to calculate the dimming potential of artificial lighting system in specifically modified version of software WILS. The regulation of lighting systems in this calculation carries out in 10 % steps ranging from 0% to 100%. For this example it has been chosen linear relationship between luminous flux and wattage. When requirements for luminous flux outgoing from the lighting system is less than 10%, the software disconnects this system. For practical calculations of lighting system saving is suitable to use a maximum of 4 independently controlled light-lines. The usage of more controlled lighting

systems (light-lines) leads to a significant increase in the length of the actual calculation

The Table 1 indicates the level of control of two lighting systems R1 and R2 at different illuminance on the outdoor unshadowed plane. For example, when the value of the horizontal illuminance level on outdoor unshadowed plane (20 klx) is the first line of the lighting system of artificial lights dims at 60% and the second line dims at 70%.

Table 1. The control of the lighting systems

Model room		
E [lx]	R1 [-]	R2 [-]
1000	1.00	1.00
5000	0.90	0.90
7000	0.90	0.90
10000	0.80	0.90
20000	0.70	0.60
25000	0.60	0.40

Table 2 compares the values that we calculated by our software with measured values. The values are in the time range from 7:00 a.m. to 3:30 p.m. and they are for February. From the Chart 1 it is obvious that the first column is the total power of the artificial lighting system which is 2.59 kWh / working hours. The second column is the values of our calculation for the uniformly overcast sky. In the third column there are real values measured from 1st December to 15th April. The table is only part of the data for a sample.

Table 2. Energy comparison of measured and calculated data

Measurement	Total consumption [kwh/day]	Consumption modelled by SW [kwh/day]	Real measured values [kwh/day]
1	2,59	1,92	2,27
10	2,59	1,94	2,35
15	2,59	1,94	1,9
20	2,59	1,81	1,71
25	2,59	1,76	2,28
30	2,59	1,74	2,49
35	2,59	1,68	1,82
40	2,59	1,66	1,33
45	2,59	1,66	1,03
50	2,59	1,61	2,31
55	2,59	1,54	1,77
60	2,59	1,52	1,3
65	2,59	1,49	1,02
70	2,59	1,46	1,05

In the Fig. 4. you can see the different wattage among the lighting system without regulation, with dimming and real values measured in the model room. When using four luminaires without dimming and control (operation during the whole working hours) there is power consumption about 2.6 kWh per working hours. When using the regulation you can see that the regulated artificial lighting system is dimmed very little in this month due to the location and the height of the Sun during the day. From the measured values it is clear that the weather conditions change. As a result the total consumption of power decreases. The graphs show trends of power consumption of irregulable lighting system, modeled lighting system by using our SW and measured real values. The result shows that the savings calculated by SW are not as big as they are achieved in real terms. As a result, it is seen that the average values are calculated by sw 1.7 kWh / day and total consumption is 120.5 kWh for 72 days calculations by comparing the measured real data 1.74 kWh / day and the total consumption of 123.4 kWh for 72 measuring days. From the results we can say that our calculations are almost identical to the real measured data.

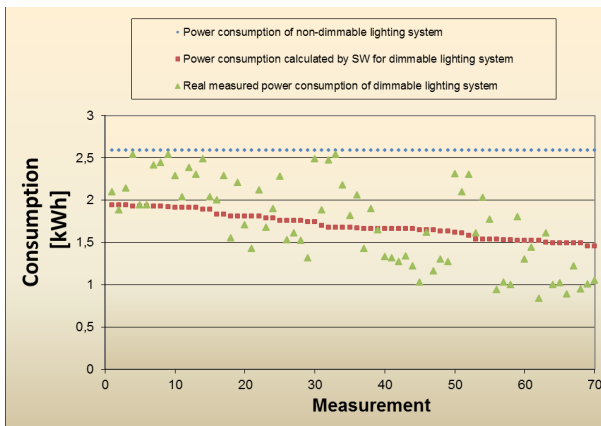


Fig. 4. The power consumption of irregulable, modelled and real measured data

3 Conclusions

In this paper was described study of smart home care designed having regard to needs of elderly people. For efficient smart home care design is necessary to start from sociological researches, which are mapping the needs of seniors. One of important items is energy power consumption in the smart home care. For modelling of possible savings in lighting systems was used software design to determine the potential savings in the combined artificial and daylight lighting systems. Its expected usage will be in the design and renovation of the lighting systems, energy audits of buildings and in recommendations of reduction of their energy consumption.

Acknowledgment

This article was supported by the project "Research of LED and OLED light sources in special applications" SP2013/88.

References

1. RYBÁR, P. a kol.: Denní osvětlení a oslunění budov, ERA 2001. ISBN 80-86517-33-0
2. DARULA, S. a kol.: Osvětlování světlovody, Grada Publishing 2009, ISBN 978-80-247-2459-1
3. Software WDLS a WILS
4. ČSN 73 0580-1 Denní osvětlení budov, ČSN 73 0580-2 Denní osvětlení obytných budov
5. ČSN EN 12464-1 Světlo a osvětlení – Osvětlení pracovních prostorů – Část 1: Vnitřní pracovní prostory
6. ŠUMPICH, J., SOKANSKÝ, K., NOVÁK, T., CARBOL, Z. Stanovení denní osvětlenosti pod rovnoměrně zataženou oblohou za účelem snížení energetické náročnosti v budovách. In: Světlo 2011. Ostrava: VŠB-TU Ostrava, 2011, s. 207-209. ISBN: 978-80-248-2480-2.
7. ŠUMPICH, J., SOKANSKÝ, K., NOVÁK, T., CARBOL, Z. Potenciál úspor elektrické energie spotřebované soustavami sdruženého osvětlení s využitím denního světla. In: Electric power engineering. Brno, 2012, s.1165-1168. ISBN 978-80-214-4514-7.
8. Gasparovsky, D., Smola, A., Janiga, P.: [Assessment of lighting systems for energy certification of buildings. PRZEGLAD ELEKTROTECHNICZNY, Volume: 84, Issue: 8, Pages: 29-33, Published: 2008, ISSN: 0033-2097](#)

Analysis of power grid including energy storage station using EMTP-ATP

Marián Uher

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
marian.uher@vsb.cz

Abstract. This article describes the quantification of negative feedback effect of renewable energy sources to the distribution system. It's all about power and voltage analysis of artificially created network, which is working under selected operation concepts as well as climatic conditions. Energy storage unit of the energy overproduced by renewable resources is then designed with regard to the incurred feedback effects.

Keywords: Power flow analysis, Smart grid, Energy storage unit, Renewable resources

1 Introduction

Renewable resources operation entails considerable difficulties in the form of feedback effecting to the operated network. It is primarily a voltage changes at the connection point, supply of reactive power into the grid, and the resulting reactive voltage drops on the line, flicker effect, bringing of higher-order harmonic voltage, etc.

Therefore, constantly rising demands on the monitoring of electrical networks and monitoring the quality of power supply. For the possibilities of feedback effects suppression is necessary to create and analyze model situations and resulting feedback effects of renewable energy sources. This requires the use of programs that utilize visualization superstructure of one of the many computational methods and algorithms.

For the analysis of these states were in an EMTP-ATP (simulation software working with the structure of FORTRAN) created a fictitious network, simulating the operation of the first Smart Grid in the Czech Republic. Therefore, this network was equipped by real renewable energy resources.

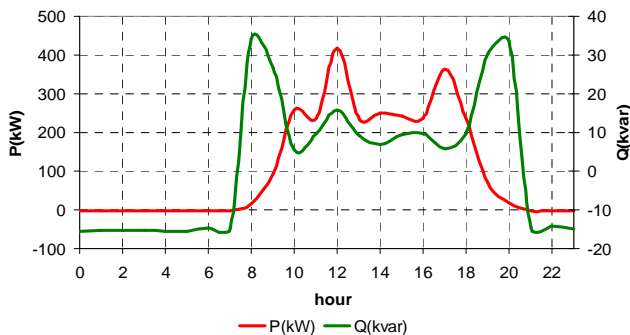


Fig. 2. PVP power supply – typically cloudy day

2.2 Wind power plant 2MW

Philosophy of creation a curve of daily power supply consisted again in the analysis of the data set of long-term measurement. The average wind speed was determined by medians analysis of data set, see Fig 3. Value of active power to this figure has been assigned from the power curve of WPP. The proportion of active and reactive power component was then set using of generator PQ diagram. WPP respects reactive power appliance.

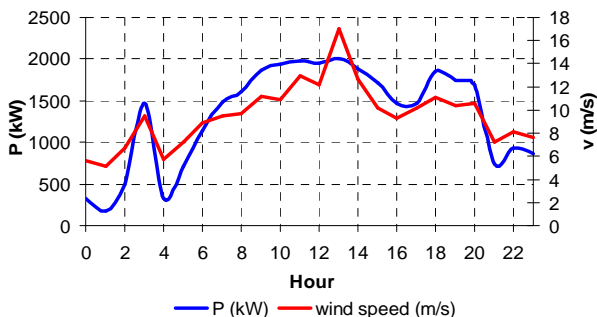


Fig. 3. WPP power supply

2.3 Cogeneration unit 1.5 MVA

Cogeneration unit has been selected as representative of power source with a constant supply of electric power. It was chosen a machine that is already installed in the future Smart region Vrchlabí. This is a four-pole synchronous generator, operating on the voltage level of 400V

2.4 Power consumption part description

Power consumption part of the analyzed network is divided into four LV power consumption units and one MV power consumption unit. LV power consumption units represent a housing development. Data set of long-term power consumption measurement of the transformer 35/10 kV was analyzed in order to match the demand also character power curve. By the corresponding class daily load diagram DLD4 (without the use of electricity for heating purposes) were determined relative power values in different time periods from the resulting curve. The resulting diagram load is recalculated using the known maximum consumed power at LV supply points.

Curve of MV power consumption was determined as an image of power consumption measured at a real industrial plant. In the ATPDraw user interface the power consumption units are presented by RLC circuits.

3 Devote power grid analysis requirements

The object of the analysis is to show the power and voltage conditions based on simulated climatic conditions and the selected operation mode of renewable resources. Demand was for thus these options:

Variant 1 - system operation with a maximum contribution of energy produced from renewable sources. Cogeneration unit assumes the active component to total power consumption of all loads. PVP operated in a typically sunny day.

Variant 2 - system operation with a maximum contribution of electricity generation from renewable sources. Cogeneration unit assumes the active component to total power consumption of all loads. PVP operated in a typical cloudy day

Variant 3 – equal balance of power active component at the transfer point (CP1). Cogeneration unit "top up" the balance of active component in the transfer point to zero PVP operated in a typically sunny day

Variant 4 - equal balance of power active component at the transfer point (CP1). Cogeneration unit "top up" the balance of active component in the transfer point to zero PVP operated in a typically cloudy day

4 Results

Large database of data was created from the values follows in each control point. The result of the analysis of power relations in CP1 of the network clearly shows the influence of non-traditional sources of the analyzed system. The largest power overflow to superior network occurs in Variant 1, where is the most visible effect of photovoltaic power plants. The overflow curve follows curve of PVP power supply.

The largest share of active power is delivered in a time of 13 hours, nearly 3 MW. From the graph there is evident the influence of WPP, which causes peak of power supply at time 3 hours 1.5 MW. The largest proportion of reactive power supplied to the superior power system is notice under the variant 2, where unfavorable weather conditions cause deterioration of PVP power factor. The highest value of reactive power is 1.4 Mvar at the time of 13:00.

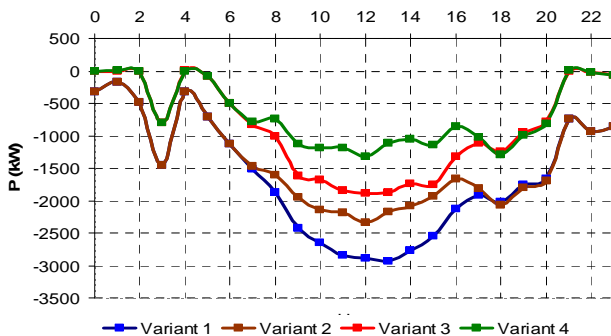


Fig. 4. Active power condition at CP1

5 Energy storage unit proposal

The energy storage unit can be considered as a measure for suppressing feedback effect. Principle leading to mitigate voltage and power changes is accumulating the energy photovoltaic peak of its power supply and subsequent delivery of power from the battery to the grid in times, when there is a lack of energy produced from renewable sources.

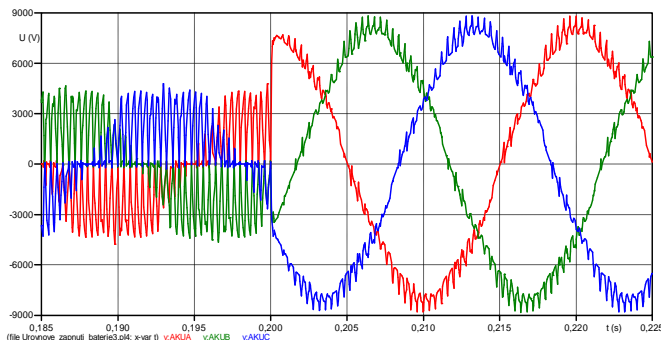


Fig. 5. Detail of the voltage in the network when connected storage unit

This storage station is in the user interface ATPDraw modeled as a battery compartment - presented by capacitor, and inverter with pulse width modulation. Estimated battery capacity is 3 MW·h, which value is assigned by practical

experiences. The inverter switching frequency is 4 kHz. Battery voltage level is 690V and works over a block transformer of 10/0.69 kV to the resources part of the analyzed system.

It can be seen inverter switching process at a time when the storage unit works to resistive load (time to 0.2 s) as well as interaction between inverter and network in time of 0.21 s. Voltage conditions at the same time are presented in Fig 5.

Conclusion

As is evident from the above text, power system operation with a local connection of renewable sources brings many risks. With regard to voltage and power conditions at the transfer point is the power system operation practically impossible without any regulation power of these resources. In this particular case, power surges and voltage variations are evident even in the case of partial power control by cogeneration units. As a measure to improve the power balance may be used energy storage system that can effectively absorb over-delivered power produced by renewable resources and prevent its overflow to the superior power system. It's also one of the requirements of existing electricity grid transition to the Smart technology. For its autonomous operation required equal balance between the energy supplied to the grid and its consumption.

References

1. Uher, M.: Optimalizace provozu sítí s obnovitelnými zdroji s využitím dynamického modelu. In sborníku, VŠB - Technická univerzita Ostrava, Ve sborníku EPE 2011, 2011, VŠB - Technická univerzita Ostrava, 978-80-248-2393-5
2. Uher, M., Mišák, S.: Smart Region Operation Optimizing Using Dynamic Model. In sborníku EPQU, Lisabon:, 2011
3. Uher, M.: Optimalizace provozu sítí s obnovitelnými zdroji s použitím Dynamického modelu, Ostrava:, VŠB - Technická univerzita Ostrava, Workshop Doktorandů 2011, 978-80-248-2378-2
4. Uher, M.: Optimizing Power Flow in the Smart Region. Ve sborníku WOFEX, Ostrava:VŠB - Technická univerzita Ostrava, 2011, 978-80-248-2449-9

Energy rating of hydrogen system at VSB-TUO

Jan Vaculík

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.vaculik@vsb.cz

Abstract. Using solar energy in photovoltaic power plants is an important method of electricity generation from renewable energy sources. Its potential is vast and technically easy to utilise. The rapid development of photovoltaic sources is having a negative effect on the electric power system control. One option for mitigating this effect is to store the energy generated by photovoltaic in times of excess power in the grid and supply it to the grid when required, i.e., during peak periods of the daily load curve. Photovoltaic power plants in connexion with a hydrogen storage system seems very promising as a source of electric power in an island (off-grid) systems. This paper describes our research into solar energy storage using hydrogen technologies in the case of off-grid systems. Storage in hydrogen as an energy carrier is currently the focus of intensive research in many research centres.

1 Introduction

Solar radiation strikes the Earth's surface unevenly and its intensity depends on the season, time of the day, and local weather conditions. As far as utilisation of solar radiation for production of electric power by means of photovoltaic power plants is concerned, the process usually occurs upon request from the linked electric power system regarding supply of power at minimum or zero level. To eliminate negative impacts on operation of electric power networks due to unsolicited supply of power, the power output needs to be limited by certain means. Should the desired solution not include non-economical disconnection of photovoltaic power plants from the network, the power generated by these as excessive at the certain moment must be stored. One of the options, which is still undergoing the research stage, deals with storage featuring hydrogen technologies.

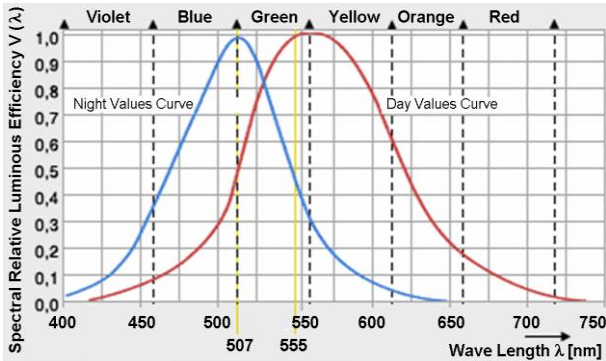


Fig. 1. Spectral curve

Besides the description of laboratory system for storage of electric power into hydrogen built in our facility, this paper deals with further analysis of data obtained through operation of the system. Attention is paid mainly to the issue of efficiency of hydrogen production using electrolysis of water and potential means for its improvement. The existing operation of this system has revealed that the weakest link used in the system is the hydrogen generator - low-temperature electrolyzer of PEM type. Research in the field of electrolytic production of hydrogen and its subsequent utilisation in fuel cells will be soon supported by launching the new laboratory of hydrogen technologies, which is being built within the Technological Centre Ostrava.

2 Solar energy

From the environmental point of view, solar energy is the most environment-friendly and cleanest source of heat and electric power. The efficiency of conversion of solar radiation into electric power per one metre of active surface reaches the level corresponding with 110 kWh of electric power per year, using contemporary photovoltaic systems. The total period of sunshine and clear skies in the territory of Czech Republic ranges between 1,400 and 1,700 hrs./year. Photovoltaic panels produce electric power even when the Sun is obscured by clouds. However, the total production represents just a fraction of actual capacity under ideal conditions. Peak output of photovoltaic panels is achieved around 12 a.m., when the radiation intensity reaches its top level. [1]

The spectrum of solar radiation is formed by the infra-red (heat), visible and ultra-violet radiation (see Fig. 1). Solar facilities help us make use of heat energy directly or convert the solar radiation energy into electric power.

2.1 Photovoltaic Power Plant Principle

Energy can be gained from the Sun directly or via indirect conversion using solar collectors or by means of direct conversion using semi-conductor photovoltaic panels.

The indirect conversion involves extraction of heat using solar collectors. The collectors are situated around the focus comprising thermal cells that serve to convert heat into electric power. Thermal-electric conversion is based on the principle of Seebeck effect. - Difference in temperature within a connection circuit with two conductors made from different material each results in generation of electric power. That is a thermal-electric cell. Characteristics and efficiency of the cell depend on properties of both metals in both conductors and the difference in temperatures between hot and cold junction respectively. An assembly of conveniently linked thermal-electric cells form a thermal-electric generator.

The direct conversion process makes use of photovoltaic effect with release of electrons within a specific substance driven by light. That happens in certain semi-conductors (silicon, germanium, cadmium sulphide, etc.). A photovoltaic cell comprises a thin mono- or poly-crystalline silicon plate, which is enriched with atoms of trivalent element (e.g. boron) on one side and atoms of pentavalent element (e.g. arsenic) on another. Once the plate is hit by photon, negative electrons will be released and "holes" with positive charge will be left behind. If both sides of the plate touch electrodes linked with a conductor, the electric current will pass through. One cm² of solar cells produces approx. 12 mA of electric current. One m² may produce up to 150 A of direct current around noon during summer days. Serial-parallel connection of cell forms a photovoltaic panel. [1]

3 Laboratory Measurement

The electric power for production of hydrogen in electrolyzer was supplied from solar panels. The block diagram below (see Fig. 2) shows linkage of individual parts of the system, including measurement points. All the power from photovoltaic panels was intended for production of hydrogen. The load on alternating current bus was not connected and the accumulator bank was fully charged.

Table 1. The parameters of photovoltaic panels

Type	SCHOTT POLY 165
Nominal power (Wp)	≥165
Voltage at nominal power (V)	35.10
Current at nominal power (A)	4.70
Open - circuit voltage (V)	43.60
Short - circuit current (A)	5.27
Module efficiency level (%)	12.60

Individual measurements were taken within specific time intervals of 0 – 60 minutes. Assessment of correct and comparable results was conducted using final values of monitored parameters only, measured within the time interval of 60 minutes.

During measurement, the electrolyzer control panel was used to change pressure of hydrogen produced within two threshold values (300 kPa - 1,379 kPa). The minimum

and maximum thresholds for hydrogen pressure were matched by measurements following the change to another vital input parameter: the temperature of reaction water. The values involved were threshold limits again, as these could be set with respect to operational safety for prevention of damage to the ion exchanger membrane inside. Measurement were taken at the temperature of 5 °C (cold water) and 45 °C (hot water).

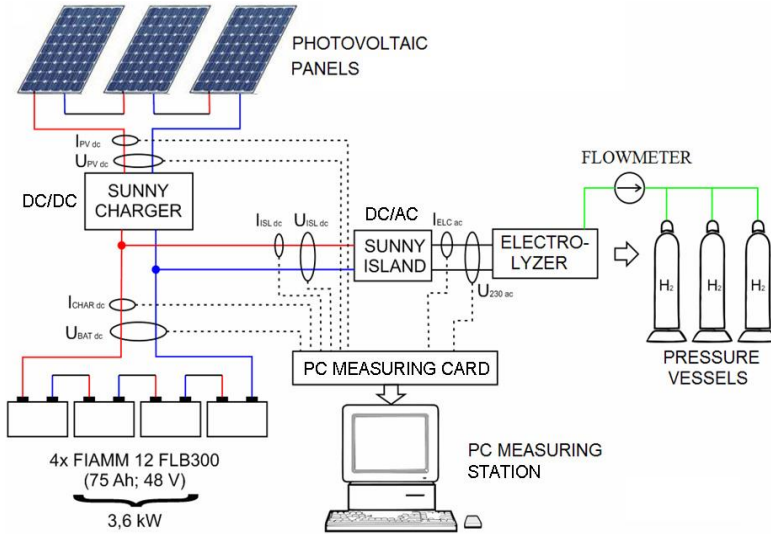


Fig. 2. Block diagram of the system measured [3]

Measurement results and values of set parameters have been included in Table 2. Figures 3, 4 and 5 show the time dependences of measured values.

Table 2. Results of laboratory measurements

Electrolyzer mode	E_{SV} (Wh)	V_{H2} (l)	E_{H2} (Wh)	Hydrogen pressure increase (kPa)	Efficiency (%)
$T_{water} = 45\text{ }^{\circ}\text{C}$, $p_{H2} = 1379\text{ kPa}$	396	40.53	119.20	28	30.10
$T_{water} = 5\text{ }^{\circ}\text{C}$, $p_{H2} = 1379\text{ kPa}$	458	40.05	117.80	21	25.72
$T_{water} = 45\text{ }^{\circ}\text{C}$, $p_{H2} = 300\text{ kPa}$	375	40.12	118.00	34	31.47
$T_{water} = 5\text{ }^{\circ}\text{C}$, $p_{H2} = 300\text{ kPa}$	397	32.41	95.32	21	24.01

E_{SV} - Electric energy consumed by the electrolyzer

V_{H2} - Amount of hydrogen produced

E_{H2} - Equivalent amount of energy in hydrogen produced

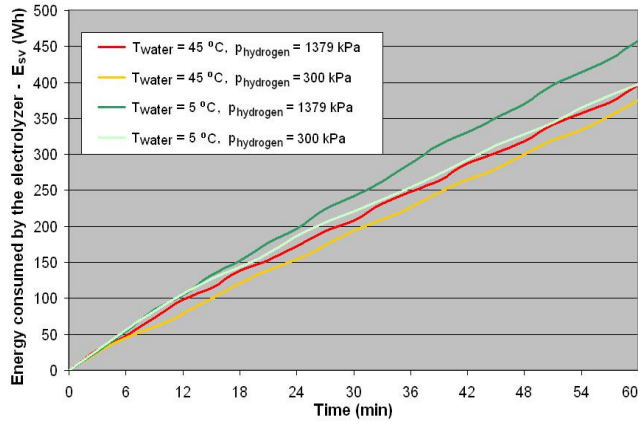


Fig. 3. Time dependence of energy consumed by the electrolyzer

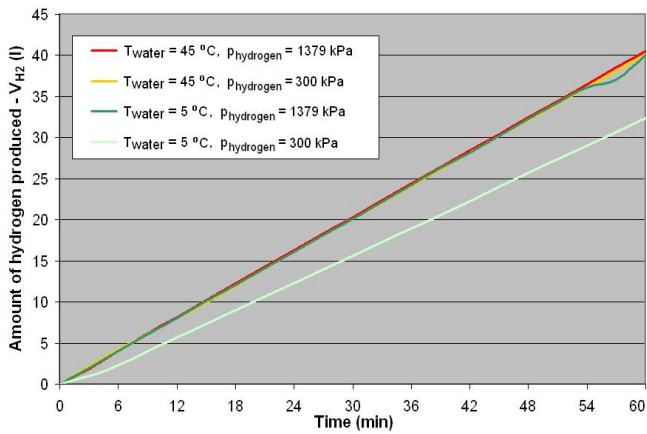


Fig. 4. Time dependence of amount of hydrogen produced

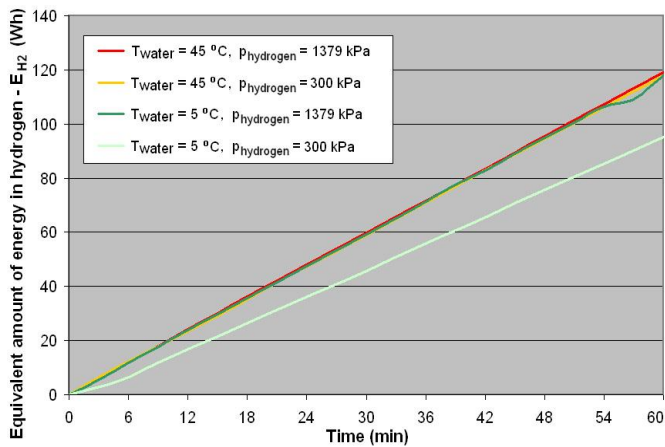


Fig. 5. Time dependence of equivalent amount of energy in hydrogen produced

4 Conclusion

The laboratory measurement described in this paper was conducted during production of hydrogen using the electrolyzer Hogen GC600 to help us determine the most convenient settings of electrolyzer parameters. The measurement was associated with a change to the vital input parameter of electrolytic production of hydrogen - the reaction water temperature. The measurement was aimed at both threshold values of potential settings for hydrogen pressure on electrolyzer, i.e. the minimum and maximum values (300 - 1,379 kPa). The main issue during this measurement was to keep the temperature of reaction water at the best constant value possible, which was achieved for both cold and hot water, with tolerance of ± 0.5 °C. The best results obtained through the measurement were based on the electrolyzer mode with minimum hydrogen pressure and hot reaction water with 375 Wh of electric power consumed and the electrolyzer efficiency calculated as 31.47 %. On the contrary, the worst results were shown by electrolyzer mode with minimum hydrogen pressure and cold water, consuming 397 Wh of electric power, which is not the highest power consumption figure, yet the electrolyzer efficiency was at its lowest (26.21 %). The above mentioned findings were then implemented as basis for operational optimisation of the whole hydrogen storage system, the most important part of which is the electrolyzer subject to examination.

ACKNOWLEDGEMENTS

This work was supported by the project ENET - Research and Development for Innovations Operational Programme No. CZ.1.05/2.1.00/03.0069, by the Czech Science Foundation - No. GAČR 102/09/1842, and by the Ministry of Education, Youth and Sports of the Czech Republic (No. SP2013/137).

References

1. Vaculik, J. : Akumulace elektrické energie získané z obnovitelných zdrojů – diplomová práce, VŠB – TU Ostrava, 2011.
2. Balajka, J.: Vodík a iné nosiče energie. Bratislava: ALFA, 1982, 303 s.
3. Hradílek, Z. - Vaculik, J. - Moldrik, P.; Konference EPE 2012. In [CD-ROM] Brno: Hotel SANTON, 2012. Storage Systems Evaluation – using Method TOPSIS MCA

Off-grid hybrid power system

Jakub Vramba

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
vramba@seznam.cz

Abstract. This paper deals with a system for power supplying energy self-sufficient house. In 2011 has been developed the power electric platform in campus of VŠB – Technical university of Ostrava. Renewable sources of electric energy belong to key areas in research and development in recent years. This course leads to the design and development of autonomous power units, which are subject to specific requirements, so-called SMART GRID system, while the specific requirements are independent of external power supply, maximum efficiency of using the renewable sources, stable generation – consumption balance and opportunities of energy storage with remote control system and monitoring. In this paper is described each part of whole system, from the sources to the storage system. Particular results of SMART GRID system were used to show the operations possibilities of this system. There are presented developmental direction as well as the area of future vision.

Keywords: Off-grid system, Batteries capacity, Voltage, Electrical energy

1 Introduction

In recent years is the emphasis on energy self-sufficient objects for different power levels, from consumption of family houses, companies up to whole cities. Energy self-sufficient is determinates as operation of family houses independently on public distribution grid. In 2011 has been established pilot project “Energy Self-sufficient Smart House Conception” to supply the family house with electrical and thermal energy using only the renewable energy sources. In this project were built up Wind Power Plant (WPP) and Photovoltaic Power Plant (PV) as an electric energy sources and solar collector as thermal energy source. In 2012 was developed comprehensive monitoring and remote control system, which is able to define basic operation parameters of each component such as whole system. For visualization of measured values was developed program in working environment LabVIEW (Laboratory Virtual Instrument Engineering Workbench), sometimes known as G-language (graphic language). This tool of virtual instrumentation is suitable for programming measuring and analyzing signals as well as for management and visualization of technological process of complicated system. This system allows enough information and data to evaluate all elements of the hybrid system. The subsequent results of this monitoring system will be used to complete the smart control system that allows intelligent controlling of whole off-grid hybrid system based on artificial intelligence.

2 The sources in Off-Grid system

The important part for power supply the family house operated in Off-grid mode is to analyze the energy consumption. The next important thing is the decision, if will be supply only the common home appliances, or even the heating system. This difference has a large effect on the installed capacity of sources in Off-grid system and battery bank. For this physical platform has been chosen a common family house, where the electric energy is not used for heating. In the family house are used common home appliances, like microwave, refrigerator and the appliance to maintain the garden. The load diagram was determined by measuring and monitoring of individual appliance for each day of the week. The measuring was carried out in several weeks to minimize the possible anomalies in the ordinary operation of the house. The basic requirements in designing Off-grid system and energy storage are:

- Weekly consumption of electric energy over 100 kW·h
- Maximum daily consumption of electric energy over 30 kW·h, average daily consumption 14 kW·h
- Maximum power peak of active power up to 6 kW

Off-Grid system is powered by two types of renewable energy sources to complement each other. The source with the biggest installed capacity is wind power plant (WPP) with apparent power of 8 kV·A. It is wind power plant with 20poles synchronous generator with permanent magnets (PMSG). WPP is equipped with a control system that operates automatically, when the wind speed exceeds the wind speed limit. The nacelle deviates from wind direction (relative angle of 30⁰, 60⁰, 90⁰), to not exceed the limit system parameters. WPP is able to operate in two modes. The first mode is public grid operation into 3phase distribution system. This mode of WPP is applied if consumption in minimal and battery bank is fully charged. The second mode of operation of WPP is power supply to Off-grid system. Variable AC Generator's voltage is applied to one 3phase rectifier, rectified DC voltage is applied to 1phase inverter that supply the Off-grid system. In this mode is whole electric power of WPP supplied to one inverter, because WPP usually works below 40 % of installed capacity. If there is a really windy day, WPP is switching from 2nd to 1st mode of operation and it allows the WPP works up to 100 % of the installed capacity. Inverters and rectifiers parameters are shown in Table 1.

Off-grid system disposes 2 strings of photovoltaic power plants. The maximum installed power of each PV is 2 kWp. The first one uses polycrystalline solar cells technology. The panels are permanently installed on the roof on the building. Polycrystalline technology of solar cells is sustainable for energy conversion of diffuse solar irradiation.

The second installation uses monocrystalline solar cell technology, which are situated on two-axis pointing device, so-called tracker. Tracker is able to increase the power production up to 30 % in direct comparison with the permanent installation. This argument is well-founded by using the database of measured values. The control unit allows maximum use of sunlight (even on cloudy days). Tracker uses to follow the sun linear motor and harmonic gearbox, which are controlled by controlling electronics. The power of both strings is connected to the 1phase solar inverter with 2

independent inputs, the parameters of solar inverter is in Table 1. Thus, the system is able to cover the consumption of 7,6 kW, when the WPP and PV inverters operate at 100 %. The wiring diagram is shown in Figure 3.

Table 1. Parameters of inverters and rectifier

3phase rectifier	P_{ACmax} kW	P_{DCmax} kW	U_{DCmax} V	η_{max} %
	6	6	600	99
1phase inverter of WPP	P_{ACmax} kW	P_{DCmax} kW	U_{ACnom} V	η_{max} %
	5.5	6	230	95
1phase inverter of PV	P_{ACmax} kW	P_{DCmax} kW	U_{ACnom} V	η_{max} %
	4.6	6	230	95

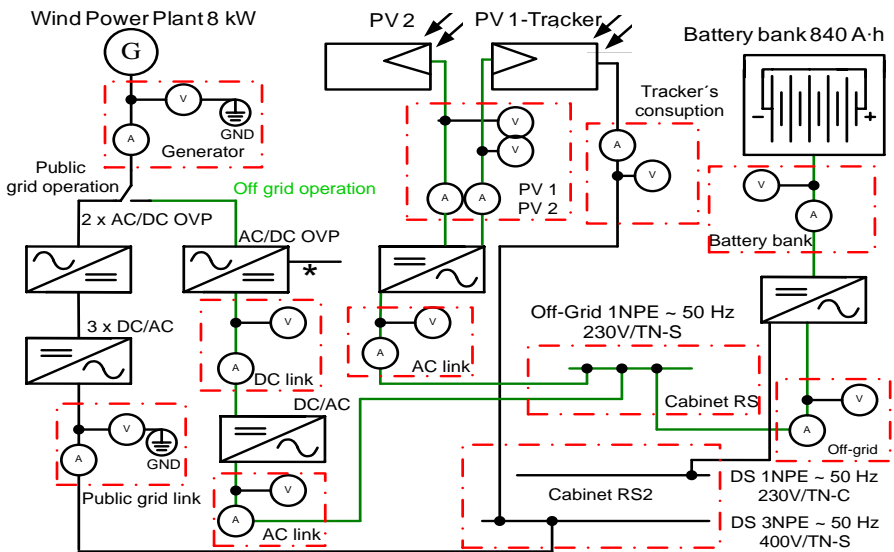


Fig. 1. Wiring diagram of family house physical platform
 * Parallel connected heating element to load resistor in DC link

3 Energy storage system

Very important in Off-grid systems is the mature selection of energy storage system – batteries capacity, this can provides enough power to cover the family house consumption and ensure minimum energy loses in generation-consumption balance. The size of the family house consumption can be contrary in the size of the power generation from photovoltaic and wind power plants.

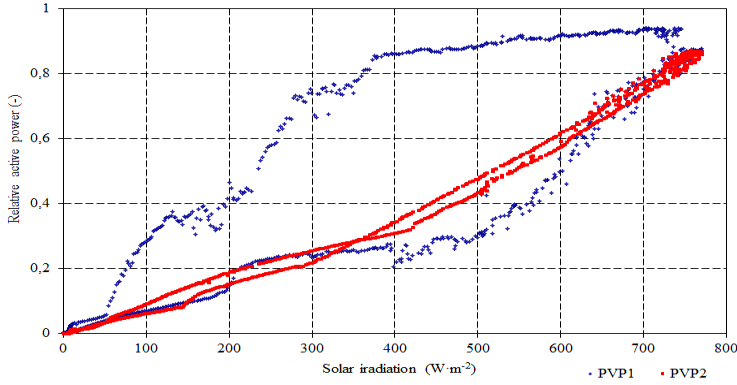


Fig. 2. Power curves of PV_{1,2}

Batteries capacity should be dimensioned enough to cover 7days of family house operation included energy losses and with no power supply. There are many types of accumulators, such as lead-acid batteries, nickel-cadmium and lithium-ion batteries, which can reach higher energy density, but their disadvantage is relatively high prices at this moment. In this system has been chosen lead-acid battery: 6OPZS. Nominal battery voltage is 2 V. Energy efficiency is about 85 % and state of charge (SoC) is 30 %. Manufacturer declares battery lifespan up to 5 100 charging cycles. Formula (1) has been used to set up the battery capacity.

$$C_B = \frac{(E_z \cdot n)}{U_B \cdot DOD_{max} \cdot T_{CF} \cdot \eta_B} (A \cdot h) \tag{1}$$

,where C_B is capacity of batteries, E_z is average weekly consumption of electric energy, n is number of days without recharging of batteries, U_B is batteries voltage, DOD_{MAX} is depth of discharge the batteries (SoC); T_{CF} is temperature correlation factor and η_B efficiency of the batteries.

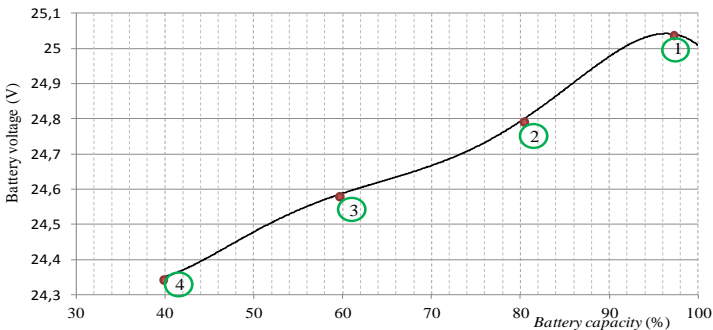


Fig. 3. Voltage and capacity of the battery bank

The batteries capacity to cover 7days consumption with losses was set up to 16000 A·h at 24 V batteries voltage level and energy efficiency 85 %. To reduce the batteries

capacity is possible to increase batteries voltage level to 48 V, then would be the capacity set up to 8000 A·h. Another possible decreasing of the capacity is replace lead-acid batteries with lithium-ion batteries at 48 V voltage level and energy efficiency of 99 % and then will be the final capacity 3200 A·h. Because of the test's regime has been set up the capacity at 840 A·h and nominal voltage level at 24 V. To manage and operate the Off-grid system it is important to know the actual capacity of battery bank. In order to determine the exact value of battery bank capacity is measuring the batteries voltage. Batteries voltage is used to calculate the actual battery bank capacity according to (1). This equation was determined by discharging the batteries by constant load with resistance character. The exact value of battery bank capacity can be determined only by discharging the batteries, because the charging voltage is higher than nominal, because of IUoU characteristics, Active inverter technology. The regression equation for calculating the current capacity of the batteries. (2)

$$C_B = \exp(-35.67 + 1.62 \cdot U_{BAT}) \quad (2)$$

, where C is battery bank capacity and U_{BAT} battery bank voltage, R_a is 0.98, what is the coefficient of determination and indicates how many percentage of dependent variables (Capacity Battery) managed to explain regression equation. This equation was created from the measured values using the software Statgraphic.

With its sophisticated battery management it is known the actually battery voltage level, shown in Figure 2 and make further system decisions based on control functions. Inverter offers a possibility of splitting the circuit in house to primary section (refrigerator, lighting) and secondary section (kettle, microwave, etc.). Figure 2 depicts the batteries SoC and the corresponding DC batteries voltage under no load. Point 1 represents 100 % capacity of batteries, thus 840 A·h, at voltage level equals to 25 V DC. Point 2 corresponds to 80 % capacity of batteries and voltage is 24.8 V DC. Point 3 depicts 60 % SoC. Voltage decreases to 24.6 V DC and central inverter disconnects secondary section in home using contactor. Point 4 is critical SoC level, batteries voltage is 24,35 V DC a batteries capacity 252 A·h. At this state limit central inverter disconnects primary section too and reconnect it back at 40% of SoC, thus at 336 A·h of batteries capacity.

4 Monitoring system

In 2012 was constructed comprehensive monitoring and remote control system, which is able to define basic operation parameters of each component such as whole system. This software was created in LabVIEW (Laboratory Virtual Instrument Engineering Workbench), sometimes known as G-language (graphic language) that uses icons instead of lines of text to create an application. This tool of virtual instrumentation is suitable for programming measuring and analyzing signals as well as for management and visualization of technological process of complicated system. Developed system is possible to divide into two parts, 1st part is monitoring and 2nd is remote control.

The 1st part monitors basic parameters of each component in long-period framework. In electric energy part are used voltage converters to measure voltage, current and subsequently the powers parameters. The measuring points have been realized systematic, see Figure 1 and measured data are processed in Figure 2. Those points allows to create the measuring chain, where is possible to observe the individual and whole hybrid system efficiencies. Another part of monitoring system is weather station, which allows evaluating of meteorological conditions. All results are saved into database for subsequently post-processing.

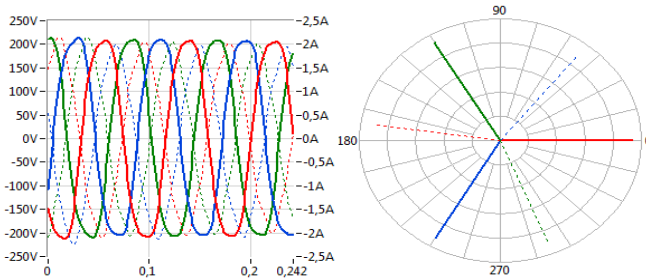


Fig. 4. Phasors diagram of wind power plant from visualization system

5 CONCLUSION

After the initial testing phase of this system and collect enough information such as overload and fault states ,the system can be used anywhere in the world such as off-grid in remote areas or in existing smart grid systems. For more information on <http://hybridni-system.vsb.cz>.

Acknowledgment

This study was conducted within the framework of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 supported by Operational Program “Research and Development for Innovations” funded by Structural Funds of the European Union and state budget of the Czech Republic, project ENET – Energy Units for Utilization of non-Traditional Energy Sources CZ.1.05/2.1.00/03.0069, Students Grant Competition project reg. no. SP2013/68, and project LE13011 Creation of a PROGRES 3 Consortium Office to Support Cross-Border Cooperation.

References

1. Andres Carvallo, John Cooper, *The Advanced Smart Grid*, Artech House Publishers, 2011, 225 pages, ISBN-13: 9781608071272
2. M. Anderson, D. Carr, Battery energy storage technologies, Proceedings of the IEEE 81 (3) (1993) 475–479

Sensorless Control of Induction Motor Using State Observers

Jiří Friedrich

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jiri.friedrich@vsb.cz

Abstract. Information about rotor position and angular speed are required for vector control of induction motor. Commonly used sensors are sources of trouble, mainly in hostile environments, and their application reduces the drive robustness. The cost of the sensors is not also negligible. All this reasons lead to development of various sensorless methods for rotor position and mechanical speed estimation in electrical drives. The paper deals with the speed estimators for applications in a structure of sensorless controlled of induction motor drive, which are based on application of Luenberger observer and so called Sliding mode observer. The mathematical description and simulation results are contained in the paper.

1 Introduction

In modern control techniques for induction motor with high dynamic requirements the speed transducer such as tachogenerator, resolver or mainly digital shaft position encoder are used to obtain speed information. These sensors are sources of trouble, mainly in hostile environments. The main reasons for the development of sensorless drives are reduction of hardware complexity and cost, increased mechanical robustness, higher reliability, working in hostile environments, lower maintenance requirements.

Removing rotor position sensors or mechanical speed sensors from a control structure of electrical drive leads to so-called sensorless electrical drive, which naturally still requires the sensors for the monitoring of stator currents and voltages. The main objective of the sensorless control is to find an estimated position of the rotor and the mechanical angular velocity which is used for the vector rotation of the variables and for a speed feedback loop.

The speed estimators, which use the motor model, can be classified into open loop estimators, MRAS (Model Reference Adaptive System) and observers (Kalman, Luenberger, Sliding mode). The deterministic observer comprises the deterministic plant model while the stochastic type (Kalman) comprises other plant model representations. In this paper are described two types of deterministic observers, the Luenberger observer and the Sliding mode observer, and analyzed those applications in the structure of vector controlled induction motor.

2 State observer

The principle of the state observer is based on the knowledge of the induction motor exact mathematical model. The basic description in the state space is defined by equation (1) and can be written using state equations (3) and (4).

$$\frac{d}{dt} \begin{bmatrix} \mathbf{i}_S^S \\ \Psi_R^S \end{bmatrix} = \begin{bmatrix} -K_R/K_L & \frac{L_h R_R}{L_R K_L} - \frac{L_h}{L_R} j\omega_r \\ L_h/T_R & -1/T_R + j\omega_r \end{bmatrix} \begin{bmatrix} \mathbf{i}_S^S \\ \Psi_R^S \end{bmatrix} + \begin{bmatrix} \mathbf{u}_S^R/K_L \\ 0 \end{bmatrix} \quad (1)$$

$$K_R = R_S + \frac{L_h^2}{L_R^2} R_R, \quad K_L = L_S - \frac{L_h^2}{L_R}, \quad T_R = \frac{L_R}{R_R} \quad (2)$$

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \quad (3)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \quad (4)$$

$$\mathbf{x} = [i_{Sx}, i_{Sy}, \Psi_{Rx}, \Psi_{Ry}]^T, \quad \mathbf{u} = [u_{Sx}, u_{Sy}]^T \quad (5)$$

$$\mathbf{A} = \begin{bmatrix} -\frac{K_R}{K_L} & 0 & \frac{L_h R_R}{L_R K_L} & \frac{L_h}{L_R} \omega_r \\ 0 & -\frac{K_R}{K_L} & -\frac{L_h}{L_R} \omega_r & \frac{L_h R_R}{L_R K_L} \\ \frac{L_h}{T_R} & 0 & -\frac{1}{T_R} & \omega_r \\ 0 & \frac{L_h}{T_R} & \omega_r & -\frac{1}{T_R} \end{bmatrix} \quad (6)$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \frac{1}{K_L} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (7)$$

where:

- i_{Sx}, i_{Sy} components of stator current space vector \mathbf{i}_S^O
- u_{Sx}, u_{Sy} components of stator voltage space vector \mathbf{u}_S^O
- Ψ_{Rx}, Ψ_{Ry} components of rotor flux space vector Ψ_R^O
- L_h magnetizing inductance
- L_S, L_R stator and rotor inductance
- R_S, R_R stator and rotor phase resistance
- T_R rotor time constant
- ω_m angular speed of rotor

2.1. Luenberger observer

Typical representative of a deterministic observer is the Luenberger observer. The mathematical principle of this method is based on the theory described in detail in [2].

Basic Luenberger observer (LO) may be used to state estimation of a time-invariant system which is described by equations (3) and (4). These equations can be modified in the form which describes LO (equations 8 and 9). [1]

$$\dot{\hat{\mathbf{x}}} = \hat{\mathbf{A}}\hat{\mathbf{x}} + \mathbf{B}\mathbf{u} + \mathbf{G}\tilde{\mathbf{y}} = \hat{\mathbf{A}}\hat{\mathbf{x}} + \mathbf{B}\mathbf{u} + \mathbf{G}[\mathbf{y} - \hat{\mathbf{y}}] \tag{8}$$

$$\hat{\mathbf{y}} = \mathbf{C}\hat{\mathbf{x}} \tag{9}$$

Where G is the observer gain matrix, which elements are discussed below. According to the above mentioned equations it is possible to draw a block diagram of the LO which is shown in Figure 1.

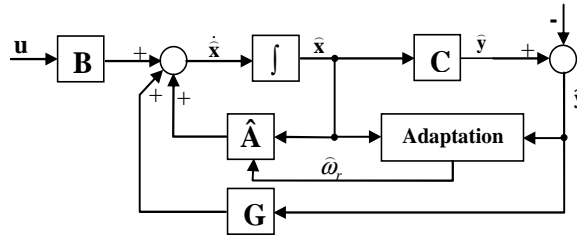


Fig. 1. Block diagram of Luenberger observer

In consideration of the above mentioned equations and the matrices of the dynamic model of the induction motor, it is possible to write the state equations for individual state quantities (10), (11).

$$\dot{\hat{\mathbf{i}}}_S = -\frac{R_S + (L_h / L_R)^2 R_R}{\sigma L_S} \hat{\mathbf{i}}_S + \frac{L_h}{\sigma L_S L_R} \left(\frac{1}{T_R} - j\hat{\omega}_r \right) \hat{\boldsymbol{\psi}}_r + \frac{1}{\sigma L_S} \mathbf{u}_S + \mathbf{G}_1 [\mathbf{i}_S - \hat{\mathbf{i}}_S] \tag{10}$$

$$\dot{\hat{\boldsymbol{\psi}}}_R = \frac{L_h}{T_R} \hat{\mathbf{i}}_S - \left(\frac{1}{T_R} - j\hat{\omega}_r \right) \hat{\boldsymbol{\psi}}_R + \mathbf{G}_2 [\mathbf{i}_S - \hat{\mathbf{i}}_S] \tag{11}$$

$$\dot{\hat{\omega}}_r = K_1 z \tag{12}$$

Where z is the correction member defined by equation (14) and K₁ is constant.

$$z = \psi_{R\alpha} (i_{S\beta} - \hat{i}_{S\beta}) - \psi_{R\beta} (i_{S\alpha} - \hat{i}_{S\alpha}) \tag{13}$$

For the adaptation algorithm, it is possible to establish two procedures, a simple integration or integration with the proportional component (equation 16).

$$\hat{\omega}_r = K_a z + K_b \int z dt \tag{14}$$

2.2. Sliding mode observer

The sliding mode theory brought to the control process new advantages, especially easier implementation and larger robustness to rotor parameter variation. The only difference in comparison to the Luenberger observer is application of function signum as it is shown in the block diagram (fig.2.) and state equation (15). This causes changes in the elements of the gain matrix G as it is shown in chapter 3.2. [1], [3].

$$\dot{\hat{\mathbf{x}}} = \hat{\mathbf{A}}\hat{\mathbf{x}} + \mathbf{B}\mathbf{u} + \mathbf{G} \operatorname{sgn} \tilde{\mathbf{y}} = \hat{\mathbf{A}}\hat{\mathbf{x}} + \mathbf{B}\mathbf{u} + \mathbf{G} \operatorname{sgn}[\mathbf{y} - \hat{\mathbf{y}}] \tag{15}$$

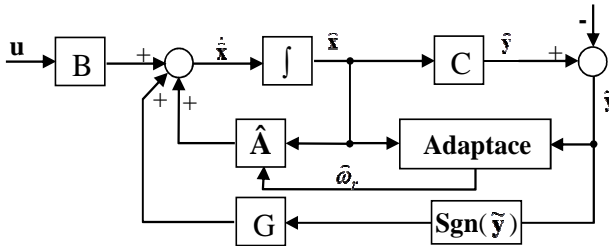


Fig. 2. Block diagram of Sliding mode observer

3 Simulation results

The behavior of the above mentioned state observers applied in the structure of vector controlled AC drive with induction motor were simulated by MATLAB-SIMULINK. The controllers in the structure of vector controlled AC drive were adjusted in accordance with the parameters of the induction motor, type P112 M04.

The simulation was realized in three speed areas to sustain appropriate behavior in large speed range. Simulation results are described lower in fig.3. to fig.6.

3.1. Luenberger observer

It was simulated Luenberger observer whose gain matrix G is defined by the conditions of proportionality and the observer poles and IM, so-called standard design matrix. First, it was necessary to establish appropriate parameters of the matrix G . Given that the reported parameters did not meet the required properties as to the accuracy of the observer had to be modified on the form. Elements of the gain matrix were following:

$$g_{1r} = 1\lambda(R_S L_R + R_R L_S), \quad g_{1i} = -10\hat{\omega}_r, \quad g_{2r} = -\frac{1(R_R L_S - 2R_S L_R)}{L_h}, \quad g_{2i} = \frac{1}{\lambda L_h} \hat{\omega}_r$$

Tuning the adaptation algorithm (by choosing the appropriate gain) was an adequate response of the observer in a wide speed range, as it is indicated in the description of the results. The adaptation algorithm gain: $K = 100$

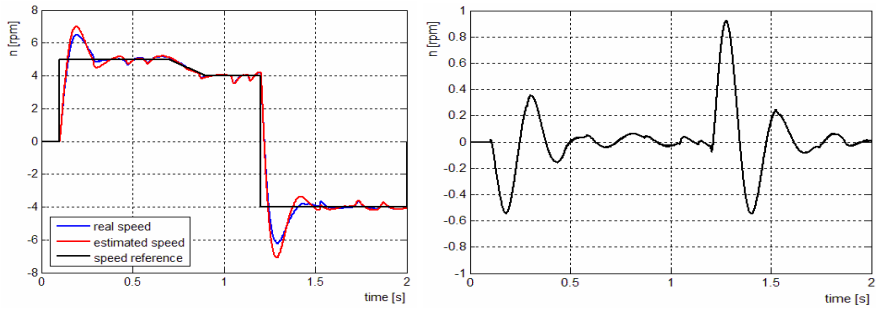


Fig. 3. Luenberger observer. Left: Reference (black), real (blue) and estimated speed (red) in area of very low speed. Right: Difference between real and estimated speed.

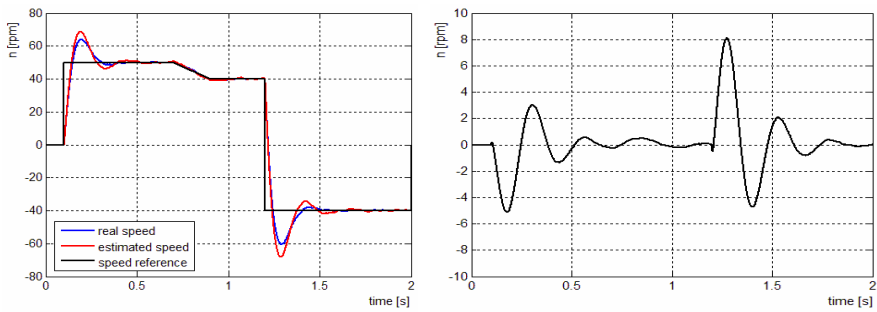


Fig. 4. Luenberger observer. Left: Reference (black), real (blue) and estimated speed (red) in area of low speed. Right: Difference between real and estimated speed.

3.2. Sliding mode observer

Appropriate elements of the matrix G were found in the same way as in case of Luenberger observer and also the tuning of the adaptation process was similar to the process mentioned in chapter 3.1. Simulation results represent fig. 5. and fig. 6.

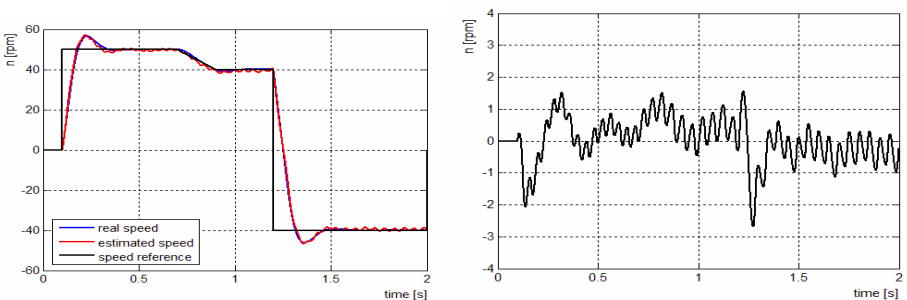


Fig. 5. Sliding mode observer. Left: Reference (black), real (blue) and estimated speed (red) in area of low speed. Right: Difference between real and estimated speed.

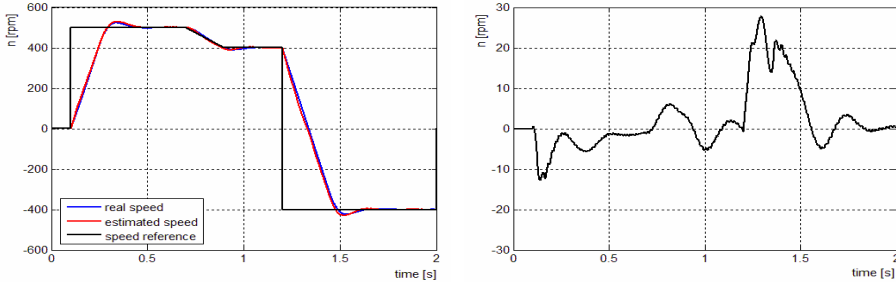


Fig. 6. Sliding mode observer. Left: Reference (black), real (blue) and estimated speed (red) in area of high speed. Right: Difference between real and estimated speed.

Conclusions

The aim of this research was to present two possible approaches to the sensorless vector control technique. It is obvious from the above presented simulation results that the application of state observers is a possible way how to replace speed encoders and get appropriate accuracy of the controlled process. The sensorless control of induction motor drive with Luenberger speed estimator gives favorable dynamic responses and the estimation of the mechanical speed is good in steady state and also in transient state. The lowest boundary of fine speed estimation is about 5 rpm, as can be seen from simulation results. In case of the Sliding mode observer, it is necessary to mention that the accuracy in area of very low speeds was not appropriate. The lowest boundary of fine speed estimation is about 20 rpm.

Acknowledgement

In the paper there are the results of the project SP2013/118 which was supported by Student Grant Competition of VSB - Technical University of Ostrava. This research was supervised by Prof. Ing. Pavel Brandštetter, CSc.

References

1. Li, H. - Xuhui, W. - Guilan, C. General Adaptive Schemes for Resistance and Speed Estimation in In-duction Motor Drives. In Computers in Power Electro-nics, 2006. COMPEL '06. IEEE Workshops on. Troy, NY, 2006, pp. 173 – 178. ISBN 0-7803-9724-X
2. Luenberger, D. G.: An Introduction to Observers. Automatic Control, IEEE Transactions on Industrial Elec-tronics, vol. 16, no. 6, pp. 596–602, 1971
3. Zhang Y. et al.: A Comparative Study of Luenberger Observer, Sliding Mode Observer and Extended Kal-man Filter for Sensorless Vector Control of Induction Motor Drives. In Energy Conversion Congress and Ex-position, San Jose, CA, pp. 2466 – 2473, 2009.

The Use of the General Theory of Electrical Machines for Modeling of Induction Motor in Matlab Simulink

Stanislav Rusnok, Pavel Sobota, and Pavel Svoboda

Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{stanislav.rusnok, pavel.sobota, pavel.svoboda1}@vsb.cz

Abstract. This article deals with the creation of a simulation model of the induction motor in Matlab - Simulink based on knowledge of general theory and mathematical description of electrical machines. Thus created model offers new possibilities and simulating more specific operating conditions.

1 Introduction

The essence of the mathematical model (general theory electrical machine) is the formulation of the basic dynamic equations for the instantaneous values essentially the same for all basic types of electrical machines. The advantage of such a description is its versatility and user-friendly way of modeling in the appropriate software. Now we will discuss a model of induction motor.

2 The mathematical description of the motor

The derivation of the differential equations is performed in the literature [1] [2]. It used the principle of transformation three-phase system into an equivalent two-phase system. In our case, we will transform differential equations into a system $\alpha, \beta, 0$. After the transformation (equations 21 and 22) now we can write the stator voltage equations:

$$u_{1\alpha} = R_1 \cdot i_{1\alpha} + \frac{d\psi_{1\alpha}}{dt} \quad (0)$$

$$u_{1\beta} = R_1 \cdot i_{1\beta} + \frac{d\psi_{1\beta}}{dt} \quad (2)$$

Rotor voltage equations can be written:

$$0 = R_2 \cdot i_{2\alpha} + \frac{d\psi_{2\alpha}}{dt} + \omega \cdot \psi_{2\beta} \quad (3)$$

$$0 = R_2 \cdot i_{2\beta} + \frac{d\psi_{2\beta}}{dt} + \omega \cdot \psi_{2\alpha} \quad (4)$$

Where R_1 is the total resistance of one phase of the stator and R_2 is the total resistance of the rotor winding calculated as the stator. Assuming that the coupled magnetic flows apply:

$$\psi_{1\alpha} = L_1 \cdot i_{1\alpha} + L_h \cdot i_{2\alpha} \quad (5)$$

$$\psi_{1\beta} = L_1 \cdot i_{1\beta} + L_h \cdot i_{2\beta} \quad (6)$$

$$\psi_{2\alpha} = L_2 \cdot i_{2\alpha} + L_h \cdot i_{1\alpha} \quad (7)$$

$$\psi_{2\beta} = L_2 \cdot i_{2\beta} + L_h \cdot i_{1\beta} \quad (8)$$

Now we substitute equations 5-8 to equations 1-4:

$$u_{1\alpha} = R_1 \cdot i_{1\alpha} + L_1 \cdot \frac{di_{1\alpha}}{dt} + L_h \cdot \frac{di_{2\alpha}}{dt} \quad (9)$$

$$u_{1\beta} = R_1 \cdot i_{1\beta} + L_1 \cdot \frac{di_{1\beta}}{dt} + L_h \cdot \frac{di_{2\beta}}{dt} \quad (10)$$

$$0 = R_2 \cdot i_{2\alpha} + L_2 \cdot \frac{di_{2\alpha}}{dt} + L_h \cdot \frac{di_{1\alpha}}{dt} + \omega \cdot (L_2 \cdot i_{2\beta} + L_h \cdot i_{1\beta}) \quad (11)$$

$$0 = R_2 \cdot i_{2\beta} + L_2 \cdot \frac{di_{2\beta}}{dt} + L_h \cdot \frac{di_{1\beta}}{dt} - \omega \cdot (L_2 \cdot i_{2\alpha} + L_h \cdot i_{1\alpha}) \quad (12)$$

Where:

$L_1 = L_h + L_{1\sigma}$ is the total inductance of the stator,

$L_2 = L_h + L_{2\sigma}$ is the total inductance of the rotor and

ω is the electrical angular velocity of the rotor.

The system of voltage equations is yet necessary to supplement the equation of motion (no load), in the form:

$$M = \frac{J}{p} \cdot \frac{d\omega}{dt} \tag{13}$$

Where J is the moment of inertia and p is the number of pole pairs. Further calculation of speed:

$$n = \frac{\omega \cdot 60}{2 \cdot \pi \cdot p} \tag{14}$$

The last equations what we need is the torque equation:

$$M = \frac{3}{2} \cdot p \cdot L_h \cdot (i_{1\beta} \cdot i_{2\alpha} - i_{1\alpha} \cdot i_{2\beta}) \tag{15}$$

After adjusting equations 9-12 on state form, we get:

$$\frac{di_{1\alpha}}{dt} = \frac{1}{L_1} \cdot \left(u_{1\alpha} - R_1 \cdot i_{1\alpha} - L_h \cdot \frac{di_{2\alpha}}{dt} \right) \tag{16}$$

$$\frac{di_{1\beta}}{dt} = \frac{1}{L_1} \cdot \left(u_{1\beta} - R_1 \cdot i_{1\beta} - L_h \cdot \frac{di_{2\beta}}{dt} \right) \tag{17}$$

$$\frac{di_{2\alpha}}{dt} = -\frac{1}{L_2} \cdot \left(R_2 \cdot i_{2\alpha} + L_h \cdot \frac{di_{1\alpha}}{dt} + \omega \cdot (L_2 \cdot i_{2\beta} + L_h \cdot i_{1\beta}) \right) \tag{18}$$

$$\frac{di_{2\beta}}{dt} = -\frac{1}{L_2} \cdot \left(R_2 \cdot i_{2\beta} + L_h \cdot \frac{di_{1\beta}}{dt} - \omega \cdot (L_2 \cdot i_{2\alpha} + L_h \cdot i_{1\alpha}) \right) \tag{19}$$

$$\frac{d\omega}{dt} = \frac{p \cdot M}{J} \tag{20}$$

Equations 13-20 is now possible using mathematical blocks gradually implement to the Matlab – Simulink program.

3 The implementation of equations in Matlab - Simulink

For greater clarity, the entire induction motor model is created using subsystems. Individual subsystems are then interconnected as to form a complete model.

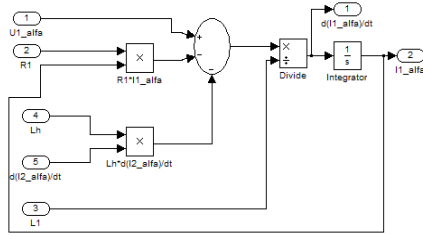


Fig. 1. Subsystem of stator currents. Equation 16

A similar way, we will proceed for the equation 17.

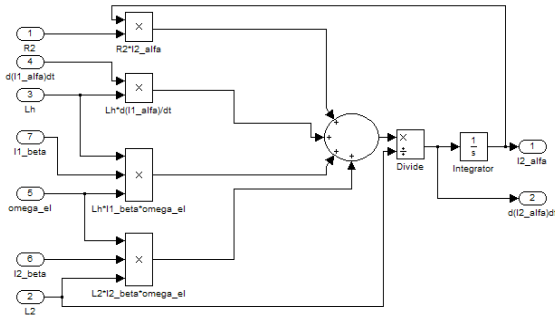


Fig. 2. Subsystem of rotor currents. Equation 18

A similar way, we will proceed for the equation 19.

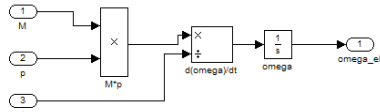


Fig.3. Subsystem of the equation of motion. Equation 20

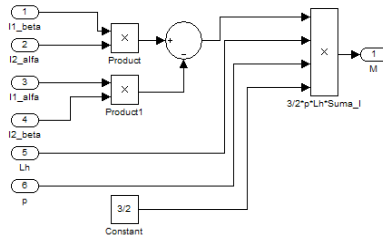


Fig.4. Torque subsystem by equation 15

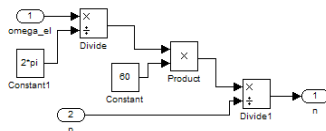


Fig. 5. Subsystem for calculating speed

To be able to use our induction motor model created in the two-phase system, we have to create a subsystem of transformation three stator voltages u_a , u_b , u_c to two stator voltages u_α and u_β by equations 21 and 22.

$$u_{1\alpha} = \frac{2}{3} \cdot u_a \cdot \cos \vartheta + \frac{2}{3} \cdot u_b \cdot \cos \left(\vartheta - \frac{2 \cdot \pi}{3} \right) + \frac{2}{3} \cdot u_c \cdot \cos \left(\vartheta + \frac{2 \cdot \pi}{3} \right) \quad (21)$$

$$u_{1\beta} = -\frac{2}{3} \cdot u_a \cdot \sin \vartheta - \frac{2}{3} \cdot u_b \cdot \sin \left(\vartheta - \frac{2 \cdot \pi}{3} \right) - \frac{2}{3} \cdot u_c \cdot \sin \left(\vartheta + \frac{2 \cdot \pi}{3} \right) \quad (22)$$

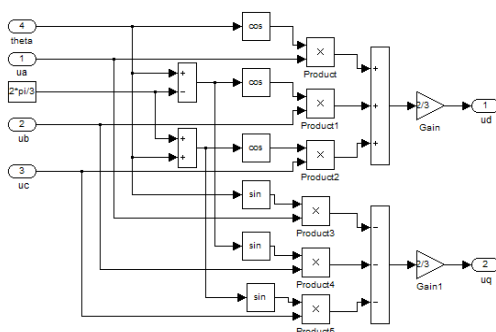


Fig. 6. Subsystem of transformation three voltages to two voltages

Now we need a subsystem for reverse transformation of two output currents to three currents. Such subsystem we create by equations 23, 24 and 25.

$$i_{1u} = i_{1\alpha} \cdot \cos \vartheta - i_{1\beta} \cdot \sin \vartheta \quad (23)$$

$$i_{1v} = i_{1\alpha} \cdot \cos \left(\vartheta - \frac{2 \cdot \pi}{3} \right) - i_{1\beta} \cdot \sin \left(\vartheta - \frac{2 \cdot \pi}{3} \right) \quad (24)$$

$$i_{1w} = i_{1\alpha} \cdot \cos \left(\vartheta + \frac{2 \cdot \pi}{3} \right) - i_{1\beta} \cdot \sin \left(\vartheta + \frac{2 \cdot \pi}{3} \right) \quad (25)$$

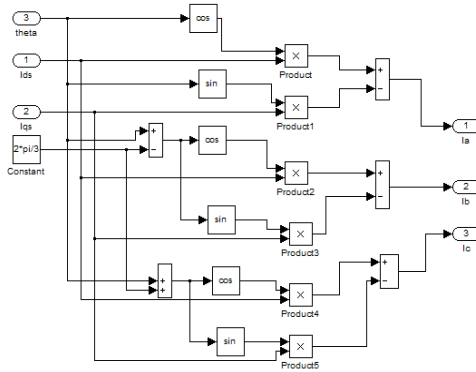


Fig.7. Subsystem of reverse transformation of two currents to three currents

Now is necessary to connect all subsystems. This creates model of induction motor.

Conclusion

The paper was given the possibility of modeling of induction motor in Matlab Simulink based on knowledge of the general theory of electrical machines. The created model eliminates the restrictions that have got models already existing in Simulink libraries. For example, it is possible to directly change size of stator and rotor resistances and the size of mutual inductance during the simulation. This makes it possible to simulate faults that occur directly in operation.

Acknowledgment

This paper was supported by project SGS SP2013/47, VŠB -TUO.

References

- [1] MĚŘIČKA, Jiří. *Obecná teorie elektrického stroje*. 1. vyd. Praha: SNTL, 1973, 162 s.
- [2] KALAŠ, Václav. JURIŠICA, Ladislav. ŽALMAN, Milan. *Technická kybernetika elektrických pohonov*. 1.vyd. Bratislava: ALFA, 1978, 392 s.

Identification and Optimization of Parameters of Synchronous Machines with PM

Ladislav Slíva

Department of Electrical Power Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
ladislav.sliva@vsb.cz

Abstract. This paper shows identification procedure selected parameters of synchronous machines with permanent magnet for optimization purposes. There is shown the steps of identification by analytical calculation, measurement, and also using the finite element method. There is also analysis of the magnetic induction in the air gap using FEM.

Keywords: synchronous machine, measurement, calculation, FEM, analysis, permanent, magnet, identification, parameters.

1 Introduction

Electrical machinery belongs to a group demanding disciplines of electrical engineering and are known are more than 100 years. Their theory was first written sometime around 1910 and so far the only rare exceptions, has not changed. Calculation methods are more precise, as well as calculations of cooling, ventilation calculations, mechanical calculations, and even machine parts. Using quality materials (particularly insulation), the performance of machines increase.

Not only at home but also abroad with both universities and private companies engaged in the improvement of various electric machines, especially the modern ones. In modern electric machine can be regarded as a synchronous machine with permanent magnet. Its use in practice gradually expanded and applied to the various electrical drives (trams, electric locomotives, etc.), among others, has a significant role in wind power as a synchronous generator with permanent magnet.

Before you start any engine optimization, you first need to understand its behavior in various configurations, and then find a possible way of optimizing. Can not only optimize the efficiency, which is very popular, but also as torque ripple or induced voltage. Very useful tool for optimization of electrical machines are used principally programs based on finite element method (FEM). Used as the 2D version, and the 3D version.

The actual work is concerned with optimization of synchronous machine with PM. In the first year of study I did identify the parameters of the replacement scheme (calculations, measurements, FEM) and also used the magnetostatic analysis in programming environments Ansys Workbench for the analysis of 3D magnetic circuit.

2 Analysis of Synchronous machine with PM

Analysis of electrical machines means to perform an analysis of its properties, ranging from analysis of parameters replacement scheme (Fig. 1). The analysis using analytical calculations, measurements, simulations and simulation programming languages, and more recently finite element method (FEM). Investigated engine is analyzed in the axis d and also in the axis q , because it has either expressed poles, or a magnet placed in the iron rotor (IPMS - Interior Permanent Magnet Synchronous Motor). Label real machines, including additional analysis of data in the following table (Tab. 1), while for obtaining the remaining data on the label usually are not trying to dismantle the machine thoroughly and measure the dimensions of magnetic circuit, including the size of the groove, count the number of wires in the groove etc.

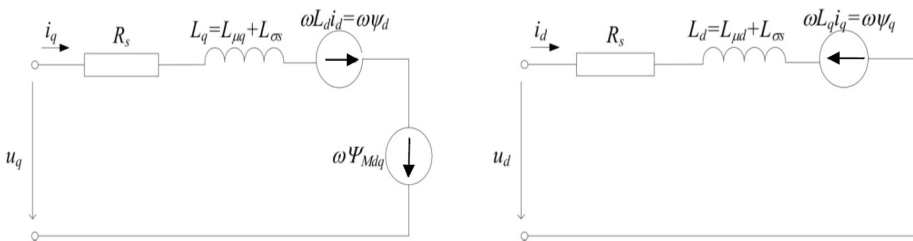


Fig. 1. Equivalent circuit IPMS with fully internally stored magnets on the axis d and q

Table 1. The label machines, including additional analysis of data identified

U_n	400 V	Rated voltage
I_n	8,3 A	Rated current
	Y	involvement
f_n	36 Hz	nominal frequency
n_n	360 1/min	Rated speed
Q_s	48	The total number of slots
Q	45	The total number of slots containing windings
q	1,25	Number of slots per pole and phase for Q
$2p$	12	number of poles
N_s	420	Number of turns of stator windings
l_i	0,14 m	Active length of iron
δ	0,8 mm	Air gap length
d	0,146 m	Outer rotor diameter
D	0,22 m	Outer diameter of stator
h_M	4,4 mm	magnet height
w_M	32,5 mm	Width of magnet
l_M	0,135 mm	Length of magnet

2.1 Analytical calculation

The analytical calculations can determine the following parameters: stator resistance R_s , leakage inductance $L_{\sigma s}$, magnetizing inductance $L_{\mu d}$ and $L_{\mu q}$, and while here I present only the final calculations. In fact, the analytical calculation is much more extensive. Were used known relations of the theory and construction of electric machines:

$$R_{s20^\circ} = \frac{l_c}{\sigma_{cu} \cdot a \cdot S_v} = \frac{205,268}{57 \cdot 10^6 \cdot 1 \cdot 0,88 \cdot 10^{-6}} = 4,07 \Omega$$

$$R_{s75^\circ} = \frac{235 + 75}{235 + 20} \cdot R_{s20^\circ} = \frac{235 + 75}{235 + 20} \cdot 4,07 = 4,956 \Omega$$

$$L_{\sigma s} = 2 \cdot \mu_0 \cdot \frac{N_s^2 \cdot l_i}{p \cdot q} \cdot \lambda = 2 \cdot 4 \cdot \pi \cdot 10^{-7} \cdot \frac{420^2 \cdot 0,14}{6 \cdot 1,25} \cdot 3,1519 = 0,026 H$$

$$L_{\mu q} = \mu_0 \cdot \frac{2 \cdot m \cdot \tau_p}{p \cdot \pi^2 \cdot \delta_{qef}} \cdot l_i \cdot k_v^2 \cdot N_s^2 = 4 \cdot \pi \cdot 10^{-7} \cdot \frac{2 \cdot 3 \cdot 0,0386}{6 \cdot \pi^2 \cdot 0,00154} \cdot 0,14 \cdot 0,936^2 \cdot 420^2 = 0,0516 H$$

$$L_{\mu d} = \mu_0 \cdot \frac{2 \cdot m \cdot \tau_p}{p \cdot \pi^2 \cdot \delta_{def}} \cdot l_i \cdot k_v^2 \cdot N_s^2 = 4 \cdot \pi \cdot 10^{-7} \cdot \frac{2 \cdot 3 \cdot 0,0386}{6 \cdot \pi^2 \cdot 0,005094} \cdot 0,14 \cdot 0,936^2 \cdot 420^2 = 0,02127 H$$

2.2 Identification of the parameters measured

For the stator winding resistance measurement is appropriate to use the VA method. Each phase was measured separately. Measurements took place at a temperature $t = 28^\circ\text{C}$. The median value of stator resistance R_s converted to 20°C is $R_{s20} = 3,93 \Omega$ and at 75°C is converted to a value $R_{s75} = 4,78 \Omega$.

Leakage reactance of stator winding $X_{\sigma s}$ (inductance $L_{\sigma s}$) can be measured using a method based on the general theory of electrical machines. Based on this theory it is possible to deduce that the non-rotating reactance $X_0 = X_{\sigma s}$ (Fig. 2).

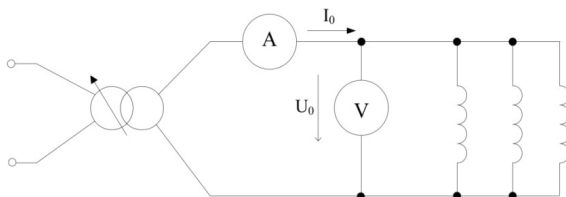


Fig. 2. Circuit diagram for measuring non-rotating reactance IPMS

During the measurement of the rotor was braked, for power supply was used single-phase autotransformer. From the measured data of voltage U_0 and current I_0 can then calculate the leakage inductance by the relation

$$L_{\sigma s} = \frac{\sqrt{\left(3 \cdot \frac{U_0}{I_0}\right)^2 - R_s^2}}{2 \cdot \pi \cdot f}$$

The following table (Tab. 2) shows the measured and calculated values, where has the motor measured stator leakage inductance $L_{\sigma s} = 0,02897$ H.

Table 2. Measured and calculated values from measurement $L_{\sigma s}$

U_0 (V)	11.468	17.187	23.24	28.67	33.23
I_0 (A)	3.355	4.995	6.78	8.459	9.93
$L_{\sigma s}$ (H)	0.028649	0.029027	0.029089	0.028794	0.028419

When measuring the magnetizing inductance following the procedure, which is proved in [1]. In primarily on instead of AC in Figure 3a join DC source. Rotor synchronous machine is, aligns "with the axis. In this position, the rotor is braked SMPM (locks) to stayed be throughout this measurement in this particular position. Then perform the measurement in figure 3a.

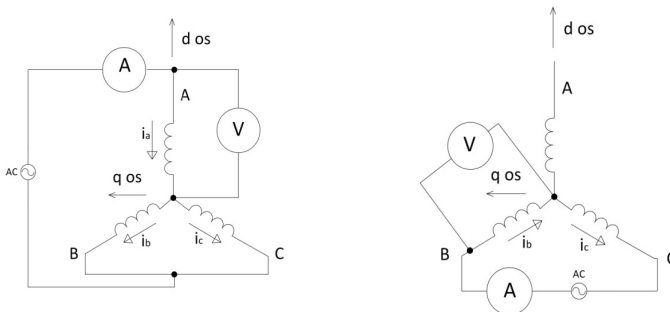


Fig. 3. Schematic diagram for measuring $L_{\mu d}$ and $L_{\mu q}$

When investigating magnetizing inductance on the axis q is needed to implement according to diagram in fig. 3b. For measured parameters are needed oscilloscope, which displays the voltage and current and measuring instruments - ammeter, voltmeter. In the first phase corresponded to the measurement diagram in fig. 3a. The measured phase A was $U_a = 142$ V and current to phase A was $I_a = 6,8$ A . The oscilloscope was then determined by the phase shift between voltage and current, which was $\Delta t = 3,88$ ms, at a frequency 50 Hz which means voltage phase $\varphi = 70^\circ$. $L_{\mu d}$ then will

$$L_{\mu d} = \frac{2}{3} \cdot \left(\frac{X_d}{2 \cdot \pi \cdot f} - L_{\sigma s} \right) = \frac{2}{3} \cdot \left(\frac{19,62}{2 \cdot \pi \cdot 50} - 0,02897 \right) = 0,02239H$$

The following connection shown in Fig. 3b. Voltage phase B was $U_b = 127,1$ V and current was $I_b = 3,95$ A. Oscilloscope to determine the shift between voltage and current that was $\Delta t = 4,5$ ms. This shift corresponds to the angle $\varphi = 81^\circ$. $L_{\mu q}$ then will

$$L_{\mu q} = \frac{2}{3} \cdot \left(\frac{X_q}{2 \cdot \pi \cdot f} - L_{\infty} \right) = \frac{2}{3} \cdot \left(\frac{32,177}{2 \cdot \pi \cdot 50} - 0,02897 \right) = 0,04905H$$

2.3 Analysis of machine parameters using FEM

Magnetizing inductance $L_{\mu d}$ and $L_{\mu q}$ can be easily analyzed using FEM. The following figures show the cross section of a synchronous machine with permanent magnets and on the axis d and q (Fig. 4a, b). Fig. 5a, b show the spatial distribution of magnetic induction B_δ in the air gap SMPM. The analysis was carried out in the FEMM 2D.

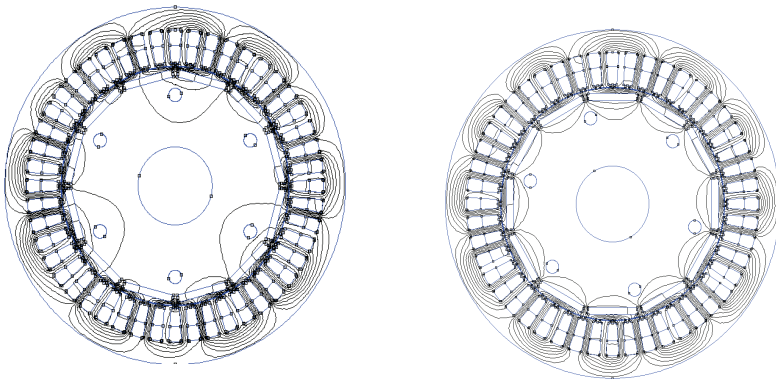


Fig. 4. Cross-section of synchronous machines with PM on the axis d and q

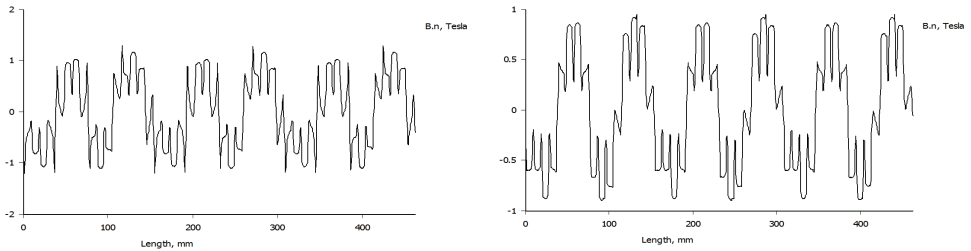


Fig. 5. Spatial distribution of magnetic induction in the middle of the air gap on the axis d and q

Subsequently, on the program Ansys Workbench was constructed 3D model and solved in idle state for the purpose of analysis of a magnetic circuit.

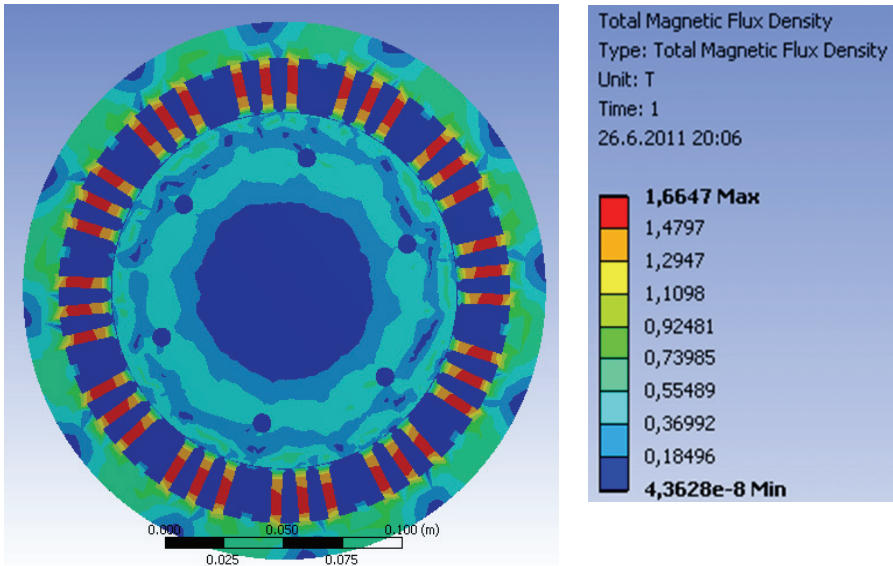


Fig. 6. Status idle (3D model)

3 Conclusion

Parameters of the measurements obtained will be used in simulations to detect transient and steady states during operation SMPM. It will also serve as a reference to compare values when creating the different topology SMPM. From 3D model is also evident that the magnetic circuit could be better utilized.

References

1. Hrabovcová, V., Rafajdus, P., Franko, M., Hudák P.: Measurements and modeling of electric machines. 2nd. ed., Zilina, 2009, ISBN 978-80-8070-924-2.
2. Gieras, F.J., Wing, M.: Permanent magnet motor technology. 2nd ed., Dekker, 2002, ISBN: 0-8247-0739-7.
3. Pyrhonen, J.; Jokinen, T.; Hrabovcová, V.: Design of Rotating Electrical Machines. 1st ed., John Wiley & Sons Ltd., 2008, ISBN: 978-0-470-69516-6.
4. Ruzojic, B. ; Zarko, D.; Ban, D.: Interior permanent-magnet motor for ship propulsion, design and testing, Power Electronics and Applications. 2009. EPE 09. ISBN: 978-1-4244-4432-8.

Sensorless Control of Induction Motor Using Injection Method

Tomáš Verner

Department of Electronics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
tomas.verner1@vsb.cz

Abstract. The paper describes the analysis of the modern way for the control of an induction motor (IM), using sensorless control using the injection methods. The paper also deals with the creation of simulation models for the sensorless control with the injecting voltage signal, which is associated to the base supply voltage. Simulation models for sensorless control are based on classical model of induction motor with the vector control. The injection is carried out in the stator reference system. The main goal for this method is control without the speed sensor. The paper presents some partial results and knowledge learned from the creation of the simulation model and his verification in a simulation tool Matlab-Simulink.

1 Introduction

Current demands for electric drives are still higher, whether it is efficiency of drive, minimizing the maximum price, or the reduction of energy consumption and maintenance costs. These reasons also relate to elimination of DC drives and exchange for the regulated AC drives. For controlling speed of the drives is essential to know the position of the rotor. The position can be obtained using a speed sensor or by using the sensorless control. Elimination of speed sensor has many advantages for a drive, such as improved reliability and the use of the drive in demanding environments, etc. Reducing costs by using sensorless control is essential for small drives, where the price of the sensor itself is comparable to the price of the drive. For a large drives sensorless control helps mainly to higher robustness.

Sensorless control gives the option to replace the speed sensor either using a mathematical model of the motor, or by detecting distortion caused by magnetic or geometric asymmetries in motors.

For the sensorless control can be used many techniques - approaches, how to evaluate the rotor position. The first methods are using estimators and observers, which are based on mathematical model of the machine. For these methods, is necessary the exact knowledge of the parameters required for the observer. For the sensorless mode are main parameters, stator resistance R_S and rotor time constant τ_R , which can be constantly estimated using "on-line" identification. For low speeds are their properties of highly unstable and almost useless. And precisely because of the instability of methods with a mathematical model, attention is turning to the use of

phenomena that do not depend on the working conditions and the drive to obtain the estimated values without limited the speed of the rotor. One possibility is to observe the magnetic or geometric asymmetry around the circumference of the rotor. For this purpose are used the injection methods, which "injected" auxiliary signal and from follow response voltages or currents are trying to get the desired information. These methods can be classified into two categories, wherein method in the first category, observes a certain geometric asymmetry caused expressed by of poles of the rotor or own eccentricity of the rotor. The asymmetry in this category is for use with asynchronous motors problematic. In the second category are included such methods, that try to follow the magnetic asymmetry, which is caused by the saturation of the magnetic circuit and this is obtained position of rotor flux, which is required for torque control, or the asymmetry originating from splined rotor, or special cage of rotor, which is designed to the spatial changes its resistance or leakage inductance. There are certain restrictions on the possibilities of using grooving to estimate the rotor position. For most structures of asynchronous motors is a proposal created for to reduce as much as possible the effect of rotor grooves, suitable ratio of stator and rotor grooves, closing rotor grooves and their skewing, so it is difficult to obtain mentioned the asymmetry.

2 Sensorless control of IM using injection method

Advantages of the sensorless control:

- (i) reduction of hardware complexity and price of the system,
- (ii) increased mechanical robustness and overall stability,
- (iii) sensorless control allows trouble free operation in dangerous and belligerent environment,
- (iv) higher reliability,
- (v) reduced maintenance requirements,
- (vi) increase noise immunity,
- (vii) unaffected moment of inertia,
- (viii) improvement of machine vibration,
- (ix) elimination of cabling for sensors, etc.

The goal of sensorless control is to find an estimate rotor position or rotor speed, which is used for the vector rotation the obtained values. If the estimated position is derived from the machine model, this is sensorless control with the mathematical model of the machine and in another case, there are methods, that do not use the model of engine and the methods that detected an asymmetry inside of the machine construction. Their research solves an area of very low to zero speed. In figure 1 shows the structure sensorless speed control of induction motor. This structure (in general) is common to all methods of sensorless control with the fact, that for each estimation method of speed and oriented values are different type of estimation blocks.

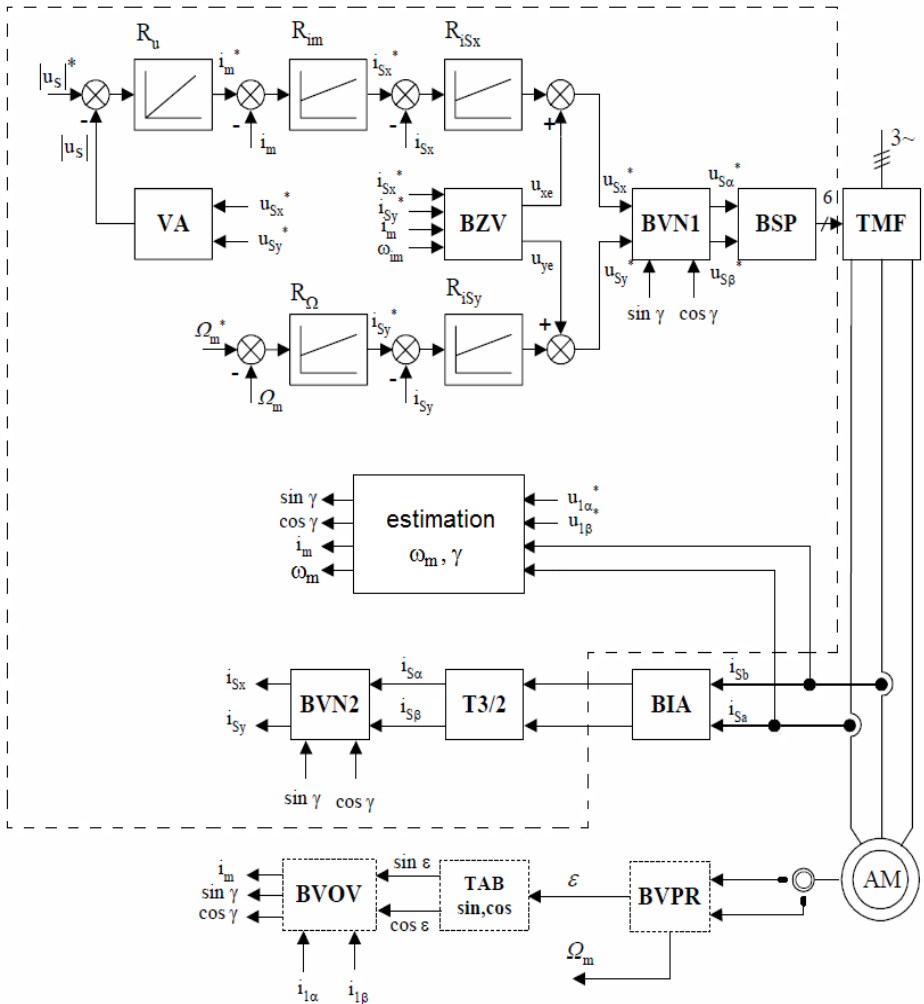


Fig.1. Control structure of sensorless vector controlled induction motor drive with mathematic model

A defined area in fig.1 is the implementation of the control system, in which for to evaluate orienting values and the need to obtain the angular speed of the rotor used block estimation. Blocks outside the defined area of the control system including an incremental encoder are in this structure included only for compare the estimated and real values and blocks are included in the structure to verify the accuracy of the evaluation of values using estimation algorithms.

2.1 Injection methods

These methods have been developed based on the requirements of applications, which require working in low and zero speed, including position control. Structures

indirectly detect some asymmetry, nonlinearity inside the machine. The estimate vector position of the rotor magnetic flux, where, with his help is estimated rotor position, can be realized by detecting the magnetic asymmetry of machines based on stator current response to injected self connected current or voltage signal with the self connected current or voltage signal with the higher frequency to stator winding. There are two main methods of injection. Injection self connected voltage or current high frequency signal or using sampling impulse. In the first method, to evocation the asymmetry is used high frequency self connected voltage signal, which is injected to stator and the signal is associate to basic voltage. This is a stable three-phase voltage with higher frequency in the range 500Hz - 2kHz. The second method is characterized by using a sampling impulse. For this method to supply voltage signal inserted measuring pulses. One of the most known methods from this group is INFORM, this method is suitable for AC machines with the geometric magnetic asymmetry.

By type of injection signal, we can divide these methods as follows:

Injection of periodic signal:

- synchronous pulsing signal injection
- high - frequency signal injection
- synchronously pulsating high - frequency signal injection
- "zero-sequence technique"

Injection of the discrete voltage impulses:

- INFORM
- injection of test impulse so called zero-sequence voltage
- injection of test impulse so called zero-sequence current

3 Method with voltage signal injection for estimation mechanical angular speed of IM

The injection takes place with the help additional voltage signal, which is superimposed to primary supply voltage (fig. 2). The frequency ranges is 500Hz - 2kHz and injection takes place in stator reference system $[\alpha, \beta]$.

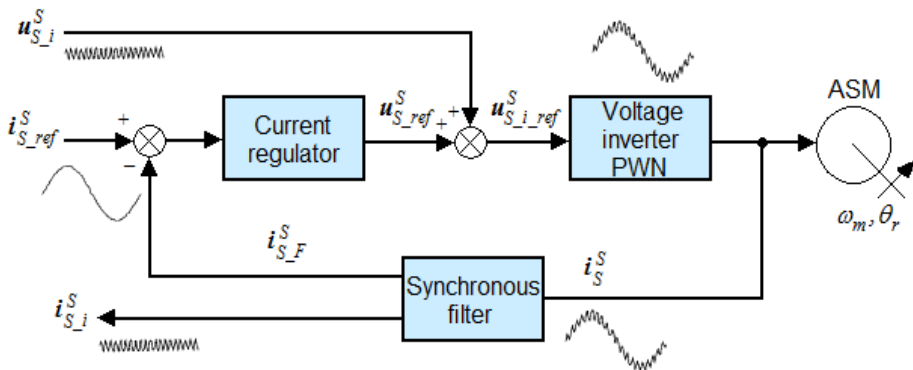


Fig.2 High - frequency voltage signal injection

Model induction motor with higher frequency, we can derive from voltage equations, which describe a model of induction motor in the stator reference frame $[\alpha, \beta]$. After all mathematical adjustments you can proceed to simulation modeling.

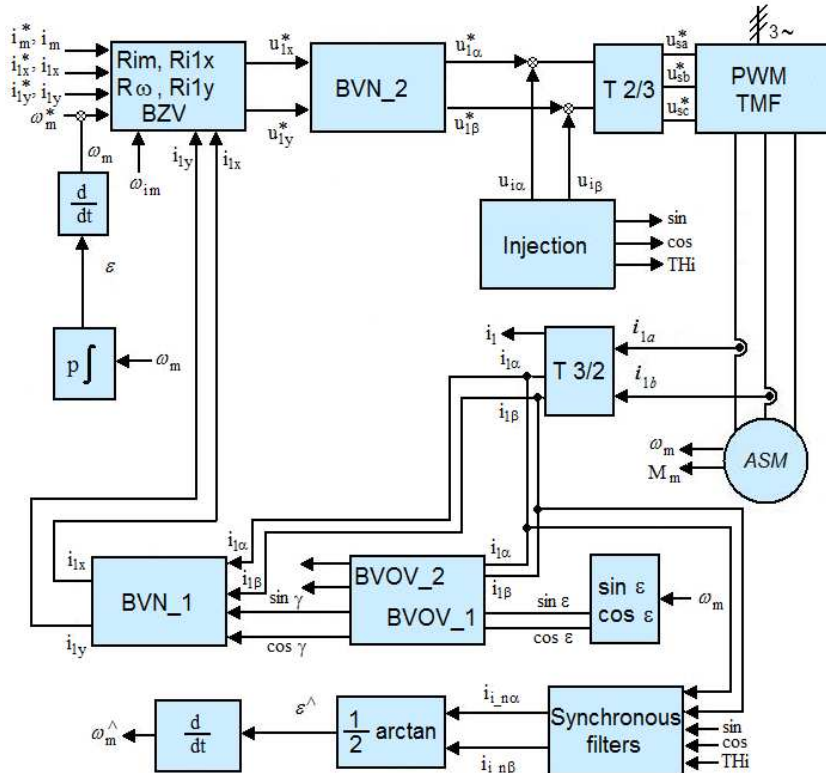


Fig.3 Simplified structure of vector control IM with the injection voltage signal

Method, with the injection voltage signal with higher frequency can be used to induction motors, which depending on the rotor position have some asymmetries. Frequency converter must be able to produce the injection signal with higher frequency.

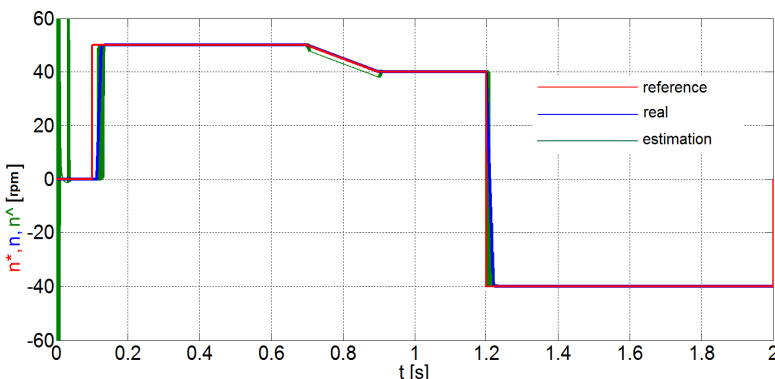


Fig.4 Reference, real and estimation speed of the induction motor

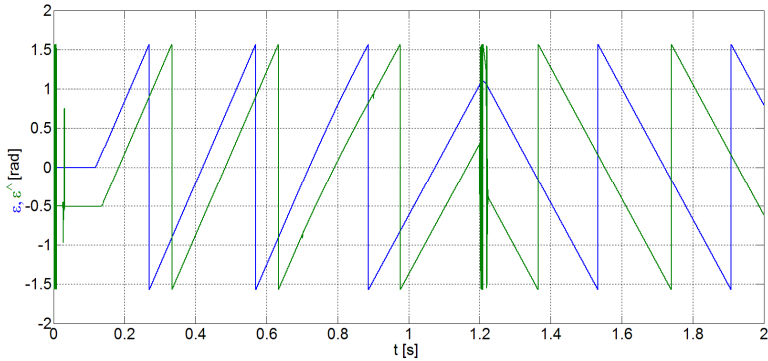


Fig.5 Real and estimation characteristic rotor angle of the induction motor

Fig.4 and fig.5 show time courses of important quantities. Fig.4 shows IM speed response and fig.5 shows rotor angle of the IM. All results and simulations models are shown in [2].

4 Conclusion

The modern control method of regulated electric drives was presented in the paper. The control algorithm uses the injection method for sensorless control of the induction motor. The injection takes place with the high – frequency voltage signal to vector control of the induction motor. There are more previews and several other possibilities of using injection methods, which aren't discussed in this paper.

Acknowledgement

In the paper, the results of the project SP2013/118 are included, which was supported by Student Grant Competition of VSB - Technical University of Ostrava. This research was supervised by Prof. Ing. Pavel Brandštetter, CSc.

References

- [1] BRANDŠTETTER, Pavel. *AC controlled drivers – modern control methods*. Ostrava, VŠB-TU Ostrava, 177p, 1999, ISBN 80-7078-668-X. (in Czech language)
- [2] VERNER, Tomáš. *Sensorless control of induction motor using injection method*. Diploma thesis, Ostrava, VŠB – TU Ostrava, 2012. (in Czech language)
- [3] HOLTZ, Joachim. *Sensorless Control of Induction Motor Drives*. In Proceedings of the IEEE, Vol. 90, No. 8, pp. 1359-1394, 2002.

Safe HMI and Car to Car Communication

Radim Hercík

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
radim.hercik@vsb.cz

Abstract. The safe HMI and Car to Car communication (SHMIC³) is a new concept to increased safety in cars and trucks. This system proposed makes use of Vehicle to Vehicle communication to retrieve and disseminate car, road and environmental conditions together with advanced Human Machine Interfaces to present the information in an optimal manner to the driver. This paper presents the research vision and challenges that will be tackled during the SHMIC³ project.

1 Introduction

Personal transportation is continuously evolving to offer a higher level of safety for people inside and outside the car. Efforts in this direction have been traditionally focused on the creation of robust mechanical structures and on reliable motion control. Additionally these safety mechanisms have been developed taking into consideration the car as a single element, without considering the cars around.

In this paper is propose an approach in which safety is developed taking actively into account the surrounding environment, cars and people. This approach exploits the possibility of radio communication between cars to broadcast different kinds of information. Ultimately part of this information is presented to a driver, hence is believe that one has to take into account how the information is presented in an intuitive manner, optimizing the response time.

2 Background

The core technologies behind the SHMIC³ system are Human Machine Interaction (mainly Heads Up Display) and Vehicle to vehicle (V2V) communication. This section gives an overview of both and sets the basis for the presentation of our new vision in Section 3.

2.1 Human Machine Interaction

Human Machine Interaction involves the study and design of how people interact with machines through machine interfaces. This discipline is an interaction between professionals with a technical knowledge such as computer engineers and professionals with competences in behavioral sciences and psychology. No complex interface can be designed correctly without the implication of professionals from both disciplines. Critical control interfaces like the ones used to acquire and present information to a driver need to be developed in conjunction with experimental psychologists and human technology engineers so drivers can “sense and interpret” faster and better. [7,8]

Some of the latest developments in car interfaces are Heads Up Display (HUD), a concept that has been based on the developments in the aviation field. A HUD projects driving information in the windshield so the driver can be looking at the road at all times. The information is projected at the appropriate field depth so the driver does not need to refocus his eye. Current research in HUDs involve optical technologies, information selection and information placement. [4]

2.7 V2V communication

Vehicle to Vehicle (V2V) communication (also known as Car-to-Car communication) is a network in which different vehicles, road signs and other elements communicate to each other in order to accomplish different goals. These goals can be at an individual level, for instance helping the driver to navigate an unknown town district, or at a higher level for example manage traffic flow in rush hours intelligently. These problems are being solved today in an individual manner without involving communication with other cars. The incorporation of V2V technologies can improve the quality of navigation and traffic management by making use of recent information providing by cars in the area. V2V can also facilitate the incorporation of other features that would benefit the traffic as a whole: increased driving safety, increased law enforcement. [1]

The developments in radio technologies conducted over the last decade (WiFi, ZigBee) and the low hardware prices have made V2V communication links feasible and affordable. However additional research effort is needed to manage the communication at a network level, to determine the information management policies and to standardized V2V solutions. [2]

3 Vision

Over the last decades research and development efforts have been focused on mechanical car protection (such as passive security improvements, Airbags, etc) leaving innovations in other fields unattended. This has been caused primarily by a design philosophy that aimed at protecting the passengers when they were having an acci-

dent instead of aiming at avoiding the accidents. Nowadays the industry is keeping protection in case of accident and developing additional preventive mechanism. These preventing mechanisms not only include mechanical systems but electrical and software systems as well. [5]

We aim at increasing safety in car transportation by exploring the application of electronic and communication technologies further. We will follow an approach in which we will start by detecting different hazards in personal transportation and we will try to prevent them by adopting and creating new technologies that a) minimize the time it takes the driver to interpret information and b) facilitate the interaction of individual cars exposed to that common hazard. We propose to focus these technological improvements in two concrete fields: the V2V field and the HMI field. We have narrowed down our research problem to the following two research questions:

- How can we use vehicle to vehicle communication to create a safer and better mean of transportation?
- What is the most effective way to present this information to the driver?

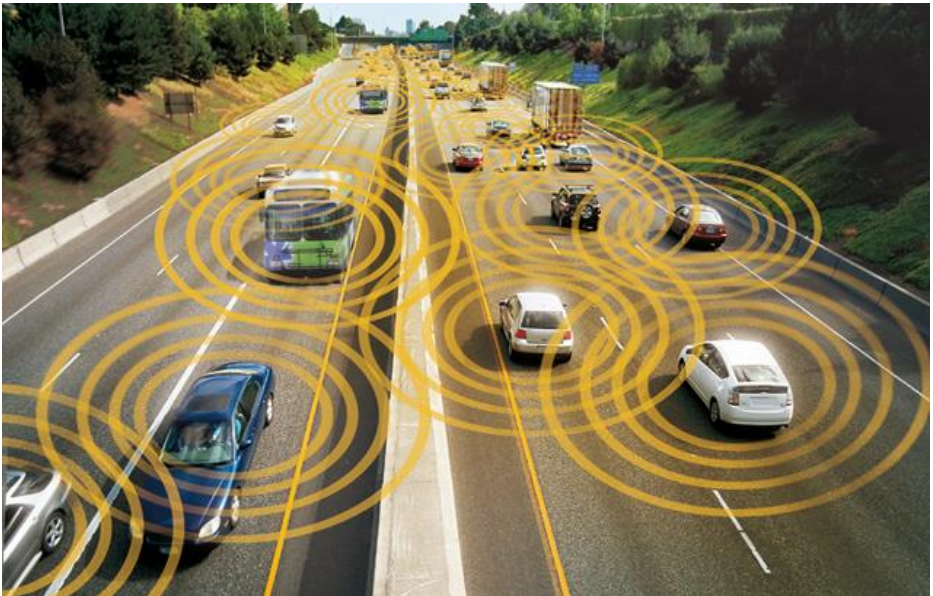


Fig. 1. Car to Car communication [9]

3.1 Research challenges

In order to answer the previous research questions we need to address a number of research challenges that we have summarized in the following lines.

Network management. A network composed by communication nodes that are constantly changing position represent a challenge from the quality of service and rout-

ing point of view. Improved protocols to target specific kind of vehicles or to target vehicles in a concrete geographical areas will be needed. [6]

Security. As a consequence of having information systems special attention has to be paid to lightweight yet reliable security protocols that ensure data and system security in terms of authentication, privacy and service availability. [3]

Systems of Systems. A system that aims at solving a problem like the one presented in this paper is complex and composed by multiple individual systems. This composition enable the set of systems to offer a more complex functionality to the user. This kind of systems are known as System of Systems (SoS). SoS engineering is a research area that is at its infancy and that it will be required to specify and study a complex emergent network of cars. [1]

Information management. The system proposed in here will need to deal with different kinds of information in a changing environment. Not all the pieces of information handled by the car system and transiting the network will have the same priority or relevance. The solution proposed in here will need to incorporate an expert system able to select what information is provided to the driver given the circumstances. [7]

4 Conclusion

This paper is focused on safe HMI and Car to Car communication and is presented a new concept approach to increased safety in cars and trucks. The new approach is based on Vehicle to Vehicle communication to retrieve and disseminate car, road and environmental conditions together with advanced Human Machine Interfaces to present the information in an optimal manner to the driver. The core technologies behind the safe HMI and Car to Car communication system are Human Machine Interaction mainly Heads Up Display and Vehicle to vehicle communication.

5 References

- [1] Possibilities of Augmented Reality in Automobile [online]. 2013 [cit. 2013-07-04]:<http://www.ocutagchallenge.com/possibilities-of-augmented-reality-in-automobile/#.UdPIWNj-L1U>
- [2] Cars turn to augmented reality [online]. 2013 [cit. 2013-07-04]:<http://www.bbc.com/future/story/20121130-cars-turn-to-augmented-reality>

- [3] Qui F., Cho J.: Automatic Identification of User Interest for Personalized Search, In Proceedings of the 15th international conference on World Wide Web, 2006
- [4] Zhang F., Song Z., Zhang H.: Web Service Based Architecture and Ontology Based User Model for Cross-System Personalization, In WI '06 Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, 2006
- [6] Abel F., Herder E., Houben G., Henze N., Krause D.: Cross-system User Modeling and Personalization on the Social Web, In User Modeling and User-Adapted Interaction, Apr. 2013
- [7] Toch E., Wang Y., Cranor L.: Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems, In User Modeling and User-Adapted Interaction, Apr. 2012
- [8] The Dozens of computers that make modern cars go (and stop) http://www.nytimes.com/2010/02/05/technology/05electronics.html?_r=1&
- [9] Emerging Connected Car Technology Tested. [online]. [cit. 2013-07-09]. Dostupné z: <http://satelliteradioplayground.com/2012/02/26/emerging-connected-car-technology-tested/>

Power Analyzer for Measurement of Efficiency of the Frequency Converter

Marek Horinek and Petr Bilik

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{marek.horinek, petr.bilik}@vsb.cz

Abstract. To measure input and output power of frequency converter in the same is quite specific task, because the output voltage frequency varies and both input and output power has to be measured from the same time window and at the same time. The integer multiply of signal cycles has to be processed. Mechanical power from analog or frequency signals from the torque and revolution sensors can be measured at the same time. The input current and output voltage contain many harmonic components and thus the sampling frequency has to be sufficiently high.

1 Introduction

The output circuits of frequency converters are made up of semiconductor switching elements such as switching transistors. The output voltage is shaped with using pulse width modulation and therefore the output voltage of converters based on this procedure contains many harmonic components that cause that the output voltage is not a purely harmonic (sinusoidal) pattern. The frequency of the output voltage is variable, it means that it can be different from AC mains frequency of 50 Hz. That is why it is not suitable for measurements which use conventional measuring devices. The standard measuring equipment designed for mains 50 Hz in this regard can only work with signals at frequencies moving in a narrow range around the nominal mains frequency, i.e. $50 \text{ Hz} \pm 10\%$. These devices are designed to measure the harmonic waveform and therefore these devices are unsuitable for accurate measurement in the electric drive system with frequency converter.

2 Hardware Platform

The power analyzer is based on cRIO hardware platform. The National Instruments cRIO platform is a modern and flexible platform consisted of several layers (see Figure 1):

- RT controller,
- FPGA (Field Programmable Gate Array),
- IO modules and communication modules,

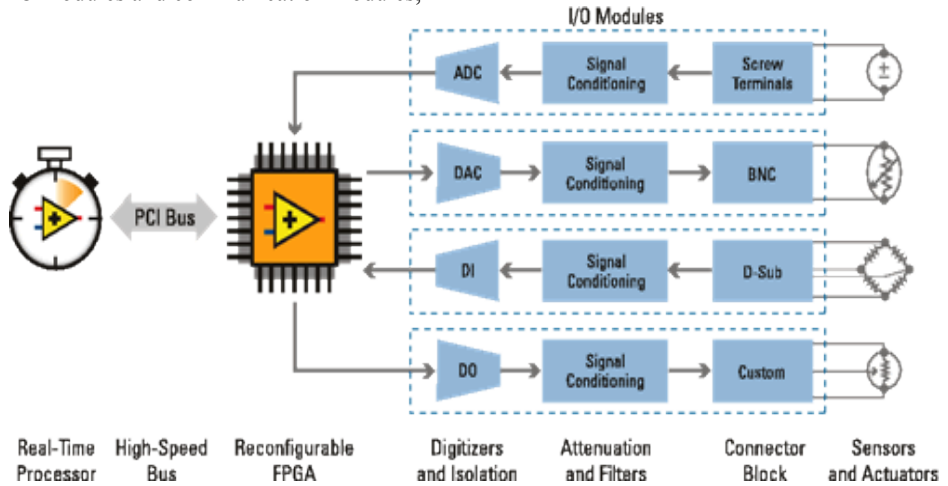


Fig. 1. Block diagram of the cRIO platform

For voltage measurements NI 9225 input modules with 24 bits A/D converter are used. The NI 9225 includes three analog inputs for direct voltage measurement up to 300 Vrms. Currents are measured indirectly using current clamps and an NI 9239. The NI 9239 module contains four analog inputs with a range of 7 Vrms. Parameters of the AD converter of this module are consistent with NI 9225.

Both input modules allow simultaneous sampling up to 50 kS/s on each channel. During sampling using these modules consistent phase shift is generated. This is important for application like this where the instant power is calculated from appropriate voltage and current samples.

Revolution sensors can be connected to the analog or digital inputs. The pulse output of revolution sensor is connected to the NI 9422 module. This module contains 8 digital inputs. Each input allows to measure signals up to 4 kHz.

3 Description of Measurement Process

The overall analysis efficiency of electric drive is based on the inclusion of the performance of individual components.

For detailed analysis of converter it can be divided into the part before DC-link and the part over DC-link. The block diagram of such system is shown in Figure 2. The mechanical power has to be measured in order overall efficiency of electrical drive can be estimated [1].

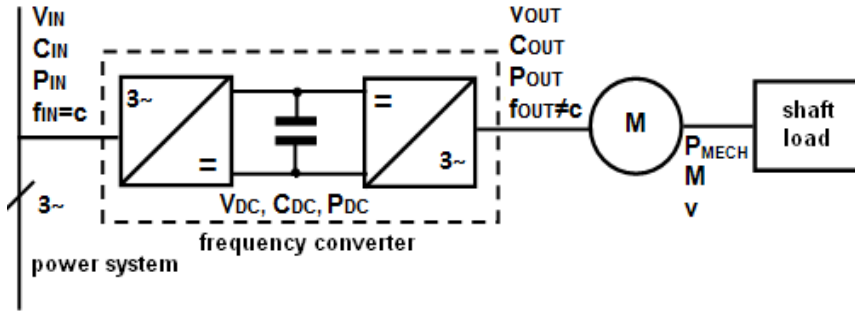


Fig. 2. The electric drive schematic

3.1 Applied Mathematical Relations

The calculation of variables is performed in a time domain. Calculations of the active power (1), RMS values (2) and (3), apparent power (4) and mechanical power (5) are done according following equations:

$$P = \frac{1}{N} \sum_{k=1}^N u_k i_k \tag{1}$$

$$U_{RMS} = \sqrt{\frac{1}{N} \cdot \sum_{k=1}^N u_k^2} \tag{2}$$

$$I_{RMS} = \sqrt{\frac{1}{N} \cdot \sum_{k=1}^N i_k^2} \tag{3}$$

$$S = U_{RMS} \cdot I_{RMS} \tag{4}$$

$$P = M \cdot \omega \tag{5}$$

3.1 Software Application of the Analyzer

The software application runs in both layers of a cRIO platform (FPGAs, RT controller) and a part of this application runs on a personal computer (Graphical User Interface - GUI).

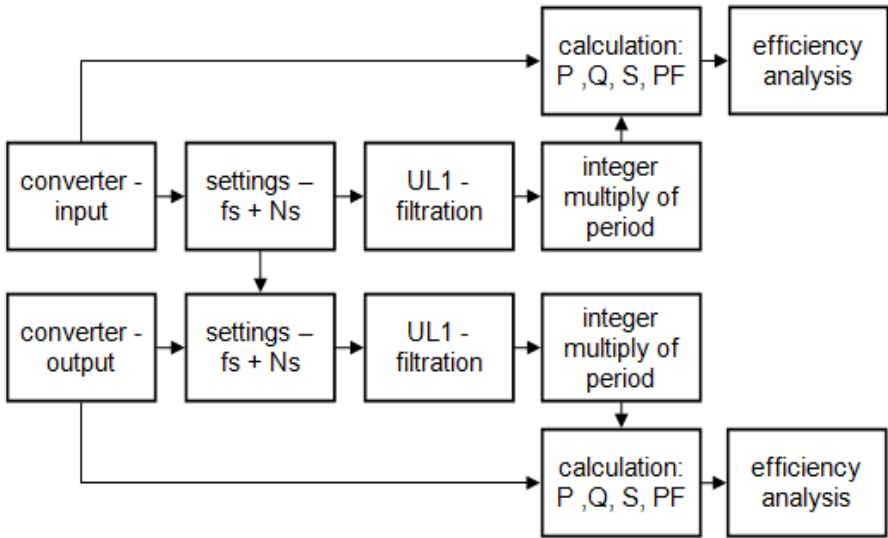


Fig. 3. Algorithm of the measurement

The lowest layer of software applications runs in FPGA. This part ensures the collection and scaling of data from analog inputs modules. From the measured values calculation of DC power is performed. In this layer measurement revolution of electric motor and torque for calculation of mechanical power is implemented.

Part of the application running in RT controller calculates the input and output powers from the measured voltages and currents. Thereafter analysis is made of the efficiency of each calculated individual powers.

To correctly evaluate the efficiency of the frequency converter from measured input and output power the following conditions have to be fulfilled:

- power has to be calculated from the same time window of measured input and output voltages and currents,
- the number of measured period signals has to be equal to the integer multiply of measured signals,
- sampling frequency should be chosen sufficiently high in order to reflect harmonic components in the spectrum of measured signals [3].

Measurement of the same time windows input and output signals is provided by an adaptive change of sampling frequency and a number of samples input and output signals. Thanks to the adaptive change of parameters described above it is ensured that the conditions for correct analyses of efficiency are fulfilled throughout the frequency range of measured signals from 10 Hz to 100 Hz.

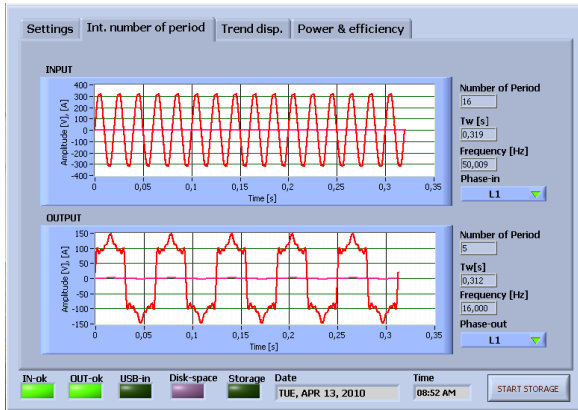


Fig. 4. Screenshot of the measurement application

Fig. 4 shows an example of measured input and output signal of the frequency converter. The input signal frequency is 50 Hz and the frequency on the output of the frequency converter is 16 Hz. Thanks to the adaptive change of the sampling frequency both input and output signals have same time window.

4 Conclusion

This paper deals with the implementation of the power analyzer for a comprehensive assessment of the efficiency of the electric drive. Measuring system is designed in the light of requirement to an accurate evaluation of converter's efficiency. These requirements include the calculation of electric power from the same time window of measured signals which consist of integer multiply of signals cycle.

The precision measurement of the power analyzer was carried out with the FLUKE 6100a calibrator. For current measurement NI 9227 modules were used in order to eliminate current sensor errors. This module is designed for direct current measurements up to 5 A, the maximum sampling frequency of the module is 50 kS / s. During testing input of the device was connected to the ac mains voltage of 230Vrms 50Hz, at the output the voltage and current were generated from calibrator FLUKE 6100a. Measurement of precision was conducted in 4 stages (sinusoidal voltage and current, voltage up to 15th harmonic sinusoidal current, voltage and current up to 15th harmonic, voltage and current up to 15th harmonic with phase shift). Amplitude of the voltage during all tests was set at 100 V and current at 1 A. All measurements were performed in the frequency range from 16 Hz to 100 Hz. According to the above mentioned tests was found the largest relative error of -0.2% of the analyzer.

References

1. BILÍK, Petr; HRUDKA, Gustav. Měření účinnosti měniče frekvence analyzátozem napájecích sítí BK-ELCOM. *ELEKTRO*. 2008, 6, s. 2.-X.
2. DUDEK, Jan: Stanoviště pro měření ztrát měničů kmitočtu, 2003 , ISBN 80-248-0106-X.
3. NOVÁK, Jaroslav; NOVÁK, Martin. Měření výkonu v obvodech s pulzně řízenými zdroji napětí. *ELEKTRO*. 2009, 1, s. 5
4. *National Instruments* [online]. c2010 [cit. 2010-03-14]. FPGAs - Under the Hood. WWW: <<http://zone.ni.com/devzone/cda/tut/p/id/6983>>.
5. *National Instruments* [online]. c2010 [cit. 2010-03-13]. NI CompactRIO Control and Acquisition System. WWW: <<http://www.ni.com/compactrio/whatis.htm>>.

E-Bike Testing Station

Jaromír Konečný

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jaromir.konecny@vsb.cz

Abstract. Quality check in industrial process is required in more and more fields. One of such field is electric bicycle. Electric bicycles become more popular for general public due comfortable and ecologic possibility to transport. Before the electric bike leaves the factory, it should be tested. This article proposes concept of testing station, which is available test electric bikes.

1 Introduction

Nowadays general people use bikes. The bikes are popular especially in Netherlands. Moreover the popularity of electric bicycles rapidly increases. [1] This article proposes concept of testing station for electric bicycles [4]. The aim of testing station is provide the complete test after production without opening the chassis or connect any cables with exception of battery. Battery pack is not part of the E-bike. The main idea is to fix the E-bike in the testing station, then simulate standard using and print certification. The E-bike support is active while pedals spinning only. The simple solution is to connect motor, which will rotate pedals and it will simulate normal operation. The powered wheel is connected to electronic brake which simulates load. During the test several quantities are measured. The E-bike testing station is able to measure pedal motor current, voltage and power. The electronic brake is able to measure, load, distance, speed etc. Moreover the battery simulator is connected and it measure input current, voltage and power. It is possible to calculate total efficiency of the bike. The concept is shown in figure 1.

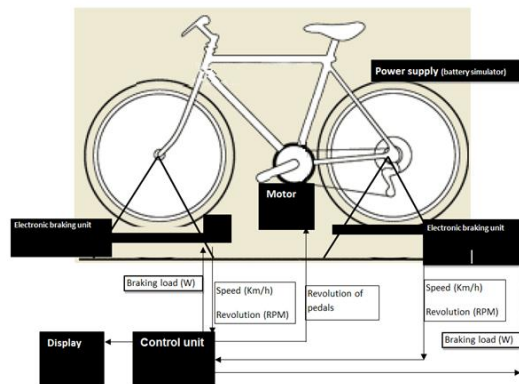


Fig. 1 Testing station concept

2 Hardware construction

E-bike testing station is constructed as modular system. Flexibility is one of the important properties. Hardware chassis is assembled from metal profiles and serve as stand. The tested E-bike is fixed into this stand. See fig. 2.



Fig. 2 E-bike testing stand

The electric part is hidden in the cupboard and it includes the whole electric hardware. See fig. 3. The heart of the stand is Main module, which is connected to the computer via USB. Main module cares for timing, measuring and data collecting from electronic brakes and measuring modules. The electronic brakes are connected wireless via Nordic chip [2]. Moreover measuring modules can be connected to testing station. In this case one measuring module is connected. Connection another measuring module is possible to connect E-bike motor and measure its power. This functionality is not suitable for testing station, because it needs open the E-bike chassis and connect measuring module between E-bike motor and E-bike control unit. The E-bike motor and pedal motor need low voltage to operate. Two power sources MANSON are available in E-bike testing station [3]. This power sources can be controlled by the main module. The pedal motor revolution is controlled that way.

3 E-bike testing

This chapter will shortly describe E-bike test procedure. One test involves several partial tests. The four basic tests are presented in this article.

1. Support test
2. Quick stop test

3. Reverse pedals spinning test
4. Load test

The first test should check the whole E-bike functionality. The pedal motor starts spinning and the support should start. The second test is quick stop test. The support should switch off in very short time maximum 280 milliseconds after the pedal motor stop. After that test the pedal motor starts spinning in inverse direction. During this test the support should not start. The last test measure E-bike's total performance. E-bike is loaded by the brake and maximum power and efficiency is determined.

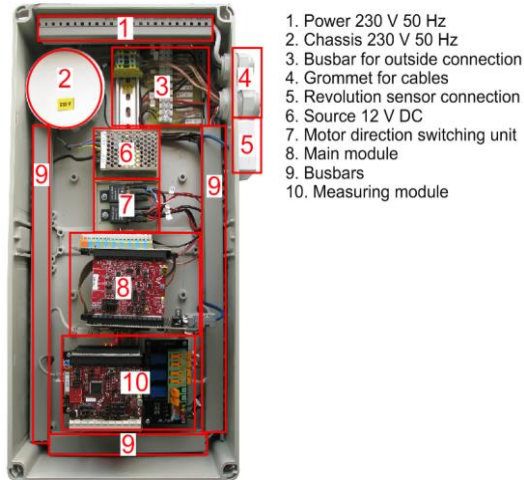


Fig. 3 E-bike testing stand electronic

One of the issues is how to determine total power and total efficiency. The E-bike control unit has not constant efficiency for different speeds. The E-bike control unit has highest efficiency when the speed and support is maximal. Consider this situation. The E-bike support is maximal and the wheel speed is also maximal. The brake load starts increasing slowly. When the brake load reaches the maximum E-bike power, the speed drops and as a consequence the control unit efficiency also drops. When the efficiency drops, then maximum power also drops. As a result the speed falls rapidly. On the other side is electronic brake. The electronic brake load is also dependent on the speed. When the speed fall down, the load also fall down. The whole system stabilizes on certain state. Unfortunately this state not corresponds to efficiency and maximum power. The measurement of the maximum power and efficiency carries out closely to point, when the speed drops.

Another issue is transient response, when the load steps up. See fig. 4. Figure shows efficiency. The blue curve shows efficiency, when the load step was 10 s. The green curve shows efficiency, when load step was 5 s. The orange curve shows efficiency, when the load changed continuously. It is clear, the fluent load change suppress the transient response.

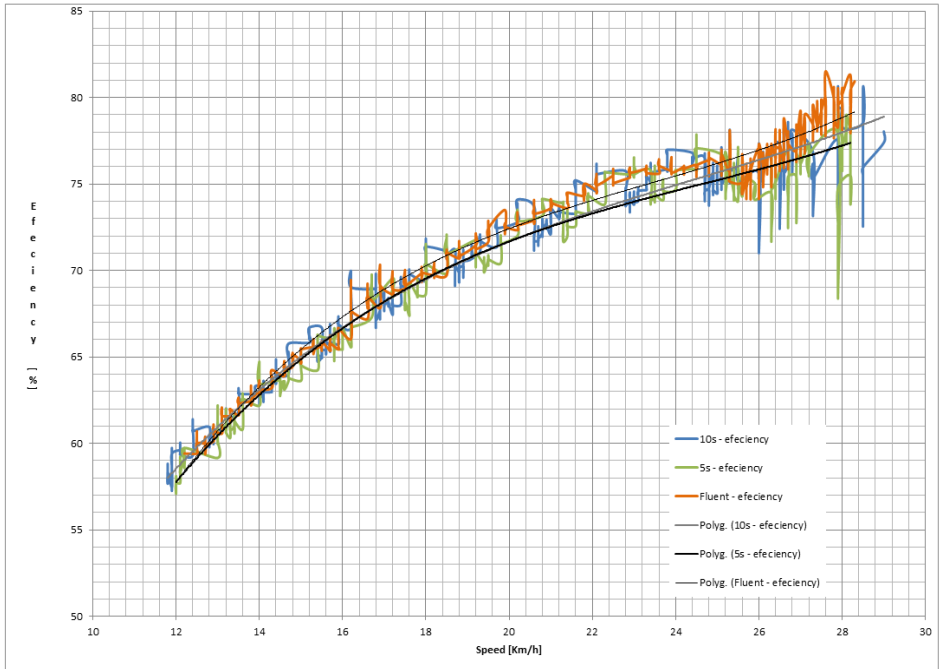


Fig. 4 E-bike efficiency test

4 Conclusion

This article presents E-bike testing station. The station is assembled in assembly line and it is used for quality check. One test takes approximately two or three minutes and it replaces test drives, which are subjective. The E-bike testing station is also able to provide sophisticate measurements. In addition, the E-bike testing station includes measurement module, which is able to measure three-phase power. The module sample frequency is 100 kHz. Moreover includes display test, battery test and any other test. E-bike testing stand provide complex measuring and testing station for quality check and research and development.

Acknowledgements

This work was supported by project SP2013/168, Methods of Acquisition and Transmission of Data in Distributed Systems of Student Grant System, VSB-TU Ostrava. Special thanks to my supervisor Jiri Koziorek.

References

- [1] Německo, Nizozemí, Čína - ráj pro elektrokola. *Hybrid.cz - auto, elektromobily, elektroskútry, elektrokola, testy - 1* [online]. 2009 [cit. 2013-06-24]. URL: <http://www.hybrid.cz/clanky/nemecko-nizozemi-cina-raj-pro-elektrokola>
- [2] *Nordic Semiconductor* [online]. 2013 [cit. 2013-06-24]. URL: <http://www.nordicsemi.com/>
- [3] *Manson Engineering Industrial Ltd* [online]. 2013 [cit. 2013-06-24]. URL: <http://www.manson.com.hk/>
- [4] *Powerflex D24 Gent: Collection E-bikes - Rock Machine* [online]. 2013 [cit. 2013-06-24]. URL: <http://www.rockmachine.us/en/e-bikes/powerflex-d24-gent.html>

Hybrid Power System Power Flow Analysis

Jakub Mañas

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jakub.manas@vsb.cz

Abstract. In conference EPE 2011 was presented the monitoring system to measure and visualize the performance parameters of the hybrid power system. Now we would like to present data analysis results computed by developed application. The application reads the measurement data (it has been acquired for more than year) from the monitoring system. This data is transformed to power flows in the individual branches, equipment's power factors and performance parameters in the graphs. The application is currently used for hybrid power system review. The paper describes achieved results and needs for further improvements of such solution.

Keywords: Hybrid Power System, Data Analysis, SW, Monitoring, SCADA

1 Introduction

A hybrid power system consisting of two photovoltaic and two wind power stations came to existence last year at VŠB-TU Ostrava. As its behaviour has to be monitored, the need for measurement and visualization at the hybrid system arose. The article describing the monitoring system was presented at Conference EPE 2011 and its brief summary is found in Part 2. Now, the hybrid system has been in operation for more than year and a lot of valuable data have been collected since then. Therefore, the application for evaluation of the measured data, which this article wants to cover, has been created.

2 The hybrid power system by VŠB-TU Ostrava

The hybrid power system consists of two photovoltaic and two wind power stations that together charge the accumulator battery. The battery is also used for the feed of two lamps of the public lighting. The block diagram of the system is in Fig. 1.

The hybrid system uses two photovoltaic panels of whose performance is 130 W (FV1) and 200 W (FV2). The transmission of the energy from the FV panels to the batteries is controlled by the regulator (REG). The wind power station VT1 uses for the conversion of wind energy to electrical energy a 200 W generator with a built-in regulator of charging, so the regulator can be connected directly to the batteries. At VT2 is, on the other hand, used a 200 W synchronous 3-phase system generator, whose performance is transferred to the battery via the rectifier and the regulator (the block with the diode, see Fig. 1). The appliances in the system are represented by two lamps of the public lighting (VO1 and VO2). VO1 is an LED lamp with the input of 40 W and

VO2 is a sodium-vapour lamp with the input of 50 W. Both lamps are operated with the voltage of 230 VAC, which is supplied there by the DC/AC converter. The battery consists of NiCad accumulators and its resulting capacity is 340 Ah at the nominal voltage of 12 VDC. [1]

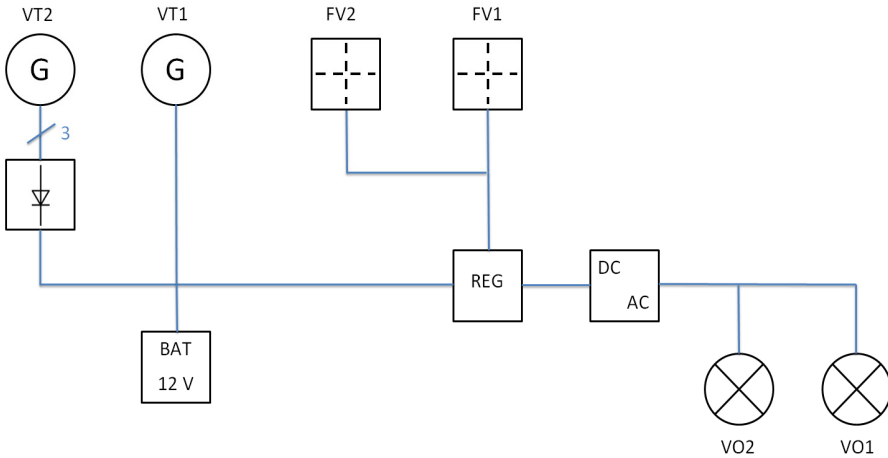


Fig. 1. The block diagram of a hybrid power system [1]

3 The monitoring system of the hybrid power system

The monitoring system is built on the HW platform NI CompactRIO. The voltage and currents in all branches of the circuit are measured, except for the appliances (the input of VO1 and VO2 is measured together). The individual measuring points are shown in Fig. 2. The monitoring system does not take any electrical energy from the batteries of the hybrid power system; its feeding is dealt with separately and provides uninterrupted operation (UPS). [1]

All the measuring HW is located in the existing outdoor distribution box where there is the whole electrical equipment of the hybrid power system. This box provides protection against weather conditions, but it is not thermally insulated. This places increased demands on the operating temperature of each component. [1]

All the components are placed on a metal structure located in the box. The dimensions of the free internal space of the box must be taken into consideration when choosing the components and the layout design. [1]

The measured data are stored in text files (daily records). The current measured data are visualized on the website by means of a web server (Fig. 3), which is operated on a remote PC. The data are also backed up on the PC and can be downloaded by authorized users with the help of the FTP access.

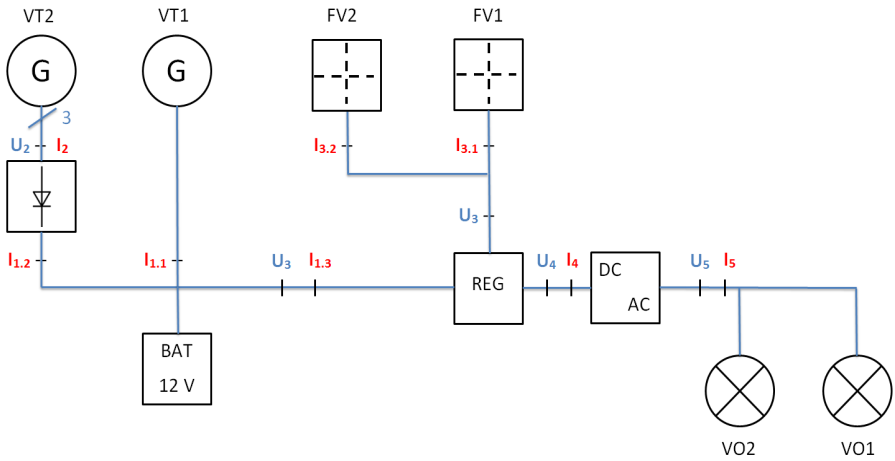


Fig. 2. The block diagram of the measured system and the measuring points [1]

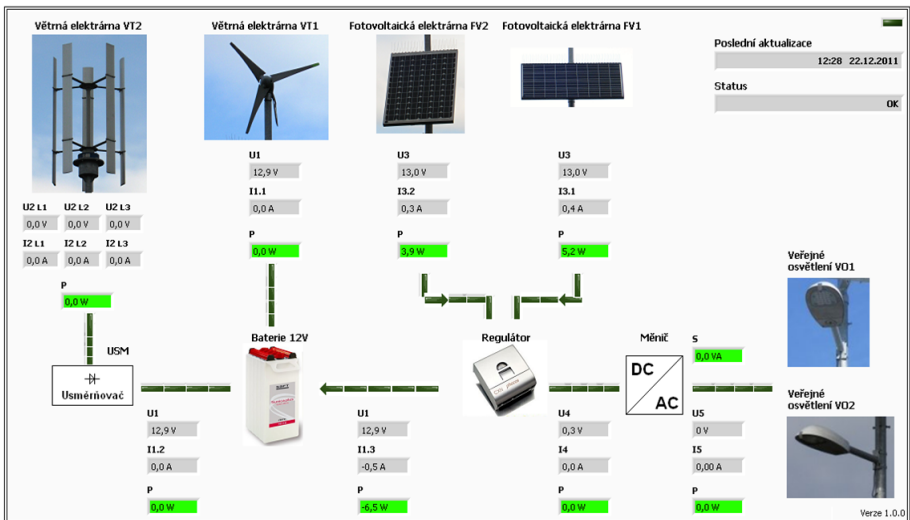


Fig. 3. Visualization of the measured data with the help of the web interface

4 Application for evaluation of the measured data

The developed application for evaluation of the measured data (application) enables graphical display (Fig. 4) of the measured and computed data (performance streams, component efficiency). The users can specify the accurate time period (the minimum resolution is 1 day) or they can choose from the menu “The last 3 days”, “The last week”, ..., “The last year” (Fig. 5). The application enables to display up to

4 parameters in the graph at the same time, when each parameter has its own vertical axis with its scaling.

The application then processes these data statistically so that the data are divided into certain time parts so that the total number of samples to the graph was approximately a thousand for each course of the parameter. The algorithm of searching for the minimum and maximum value of the parameter and their display together with their time in the correct order is the reduction of the data.

The displayed data can be exported in the form of a picture, MS Excel table, text file and TDMS file (for data processing in the application NI DIAdem).

In a simplified way, developed application is data reader, so its accuracy and reliability are related to the monitoring system. If data files are corrupted (for example unexpended restart of the monitoring system), developed application ignores the corrupted data blocks. If the sought-after parameter is not found in the measuring files and is not located in other files, this day is filled in the data with an invalid value that is not displayed in the graph. In case of low current amplitudes, power evaluation is inaccurate, so users can set power thresholds in application settings to ignore the inaccurate results. [2]

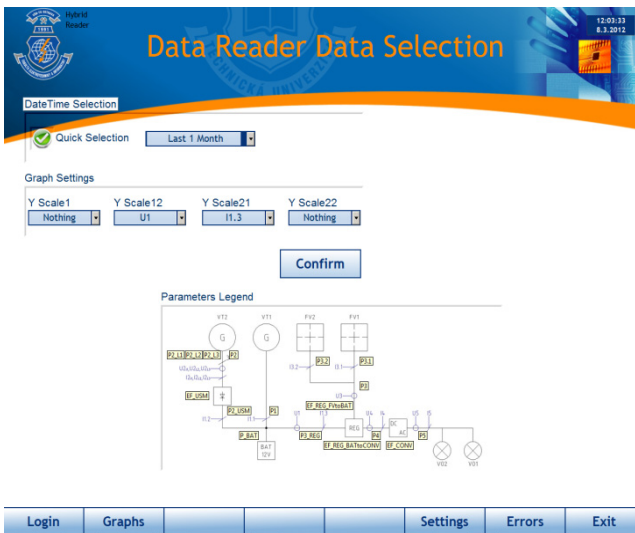


Fig. 4. GUI of the application created for the evaluated data selection

5 Conclusion

The goal of this article was to describe the developed application for evaluation of the measured data from the hybrid power system consisting of two photovoltaic and two wind power stations.

The application allows an easier and more efficient analysis of phenomena in the hybrid system at VŠB-TU Ostrava. When the application NI DIAdem for the analysis of the measured data is used, the compilation of measurement protocols will be faster

as well. These significant time savings enable the users to concentrate their time and energy on substantial issues during the research of the hybrid system.

Future development of the application will focus to create algorithms for more complex data analysis in the hybrid system.

The application uses the concept of virtual instrumentation, so it is expandable according to the future requirements without the need of higher time sources.

Acknowledgment

This report was prepared thanks to financial support of the project SGS no. SP2012/144 VŠB-TU Ostrava, FEI.

References

1. KVAPIL, Jakub; BILÍK, Petr; MIŠÁK, Stanislav: *Monitoring system of the hybrid power system* (Monitorovací systém hybridního napájecího systému). In Proceeding of the 12th International Scientific Conference Electric Power Engineering 2011. Kouty nad Desnou : VSB- Technical University of Ostrava, 2011. s. 1-3. ISBN 978-80-248-2393-5.
2. *National Instruments - Test and Measurement* [online]. 2010 [cit. 2012-05-03]. Dostupné z WWW: www.ni.com.

Knowledge-based Monitoring and Re-adaptation of PID Controllers

Jana Nowaková and Miroslav Pokorný

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jana.nowakova, miroslav.pokorny}@vsb.cz

Abstract. The classic monitoring systems are widely used in industry. The monitoring system which are based on non-conventional methods for deciding, take into account the non-numeric terms and they are open for adding more rules, are not so common. Presented system is determined for systems of second order and it is based on two expert systems. The first one for monitoring, which is performed by the fuzzy expert system of Mamdani type with two inputs - settling time compared with the previous settling time (relative settling time) and overshoot. The second one for following re-adaptation of classical PID controller. The proof of efficiency of the proposed method and a numerical experiment is presented by the simulation in the software environment Matlab-Simulink.

Key words: Expert system, PID controller, monitoring, re-adaptation, Ziegler-Nichols' combined design methods, fuzzy system, negative feedback.

1 Introduction

The paper is focused on adjustment of the monitoring system for deciding when re-adapt the classic PID controller [1], [2]. It is created fuzzy expert system of Mamdani type (ES1) with two inputs - overshoot and relative settling time and one output - score. The score determines if it is necessary to re-adapt the controller.

The following design of parameters of classic PID controller is done by the second fuzzy expert (ES2) system with a knowledge base is built on know-how obtained from the combination of the frequency response method and the step response Ziegler-Nichols design method [3].

2 Monitoring System

The monitoring system is fuzzy expert system [5], [6], [7] of Mamdani type with two inputs, knowledge base with linguistic rules and one output and it has been

created and the efficiency is proofed for controlled systems of the second order with transfer function in the form

$$G_S(s) = \frac{1}{a_2s^2 + a_1s + a_0}. \tag{1}$$

2.1 Inputs - Relative Overshoot and Relative Settling Time

The first input is the linguistic variable relative overshoot (RO) - the difference between the controlled value (CV) and the required value (RV) is rated relatively to the required value (2)

$$RO = \frac{|CV - RV|}{RV}. \tag{2}$$

The second input is the linguistic variable relative settling time (RST). As the name says, it is not the classic settling time (ST_k), it is defined as the part or multiple of previous settling time ($ST_k - 1$)(3)

$$RST_k = \frac{ST_k}{ST_{k-1}}. \tag{3}$$

For evaluation of the settling time the 3 % standard deviation from steady-state value [8]. The linguistic values of both linguistic variables are expressed by fuzzy sets, for each linguistic variable by three linguistic values (Figure 1, 2).

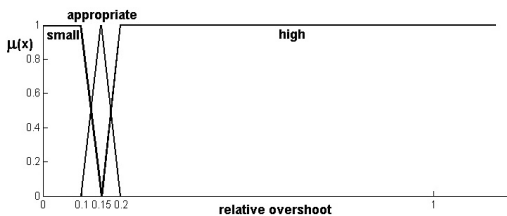


Fig. 1. The shape of the membership functions of linguistic values for input linguistic variable Relative Overshoot (RO)

2.2 Output - Score

The output of the monitoring fuzzy expert system is the score (Figure 3). As the fuzzy expert system of Mamdani type is used [9], the linguistic variable score must be defuzzificated - the COA method (Center of Area) is used [10].

2.3 Knowledge Base

The knowledge base is formed by nine linguistic IF-THEN rules of the Mamdani type [10]. The shape of membership function of output variable is inferred using the Mamdani method. The crisp value of the output score is determined using the defuzzification method Center of Area [10].

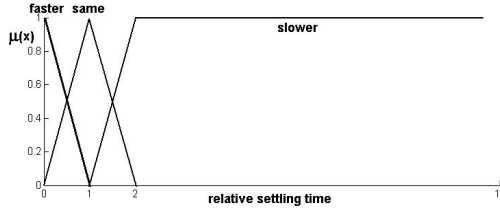


Fig. 2. The shape of the membership functions of linguistic values for input linguistic variable Relative Settling Time (RST)

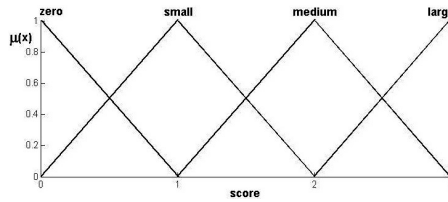


Fig. 3. The shape of the membership functions of linguistic values for output linguistic variable Score

3 PID Parameter Design System

As it was mentioned also for design parameters of conventional PID controller fuzzy expert system (ES2) is used [3], [4]. It uses know-how obtained from the combination of the Ziegler-Nichols’ design methods. [8], [11]

The constant a_2 , a_1 and a_0 from the denominator of the transfer function of the controlled system (1) are the inputs of the expert design systems (ES2).

The outputs $KKNOW$, $TIKNOW$ and $TDKNOW$ of ES2, which are also constants, are the parameters of conventional PID controller with the transfer function expressed as

$$G_R(s) = KKNOW \left(1 + \frac{1}{TIKNOW \cdot s} + TDKNOW \cdot s \right). \tag{4}$$

Expert design system is model of Takagi-Sugeno type [9] so it does not require defuzzification. For detailed information see [3].

4 The Description of Implemented Algorithm

It is important to define the algorithm of monitoring and following re-adaptation of the controller. The relative overshoot is monitored and stored in memory after every step change of controlled value. The settling time is measured also after every step change of controlled value and is assessed to the previous settling time. If these two monitored parameters are obtained the score is assessed (section 2.2).

$$Score \left\{ \begin{array}{ll} \geq 2 & \text{do not re-adapt} \\ > 2 & \text{re-adapt} \end{array} \right\}. \tag{5}$$

The re-adaptation (change of parameters of controller) procedure ES2 is done only after the change of required value.

5 Verification of Created System

The verification was done in Matlab-Simulink [12], the timing with description of important moments is depicted in Figure 4. Verification of the re-adaptation procedure proposed above is started using the controlled system (S1) with transfer function

$$G_{S1}(s) = \frac{1}{2s^2 + 9s + 7} \tag{6}$$

for which the controller with transfer function

$$G_{R1}(s) = 6.1 \left(1 + \frac{1}{0.58s} + 0.14s \right) \tag{7}$$

designed through the identification system ES2 is used.

At the time t_A the unit step of required value is introduced. Therefore, the control process is carried out with 14%-overshoot and settling time $t_{st1} = 4.3 \text{ sec}$. The appropriate calculated score by ES1 is

$$score_1 = 2.10 > 2, \tag{8}$$

which corresponds to the satisfactory control course.

At the time t_C a sudden change of the controlled system (to controlled system S2) is simulated from the transfer function $G_{S1}(s)$ to the transfer function

$$G_{S2}(s) = \frac{1}{16s^2 + 18s + 15}. \tag{9}$$

Thus, a non-zero control deviation appeared which is compensated by the original controller $G_{R1}(s)$. The oscillating control course appeared with 12%-overshoot and settling time $t_{st2} = 8.4 \text{ sec}$. Now, the calculated appropriate score is

$$score_2 = 1.66 < 2 \tag{10}$$

and insufficient control course is now indicated.

Therefore, when the nearest change of the deviation appeared at the time t_E (the unit step of required value is introduced) the re-adaptive process ES2 is initialized and it is designed a new controller with transfer function

$$G_{R1}(s) = 5.1 \left(1 + \frac{1}{0.80s} + 0.20s \right). \tag{11}$$

Now, the control process is carried out without any overshoot and with settling time $t_{st3} = 5.5 \text{ sec}$. The calculated appropriate score value is

$$score_3 = 2.65 > 2 \tag{12}$$

and the satisfactory control course is restored again.

6 Conclusion

The procedures of control quality monitoring and necessary re-adaptation of PID controller is solved using the fuzzy-logic principle through the rule-based expert systems. The first one concludes the initial impulse for controller adaptation. The rule base is created within two input linguistic variables - namely the relative settling time and relative overshoot are mentioned. The following design of parameters of classic PID controller is done by the second fuzzy expert system with a knowledge base which is built on know-how obtained from the combination of the frequency response method and the step response Ziegler-Nichols design method. The proof of efficiency was done using simulations in Matlab-Simulink. It is shown that presented monitoring system with following design of PID controller is useful for family of controlled systems of second order. The both described knowledge-based systems are open. Therefore, next time authors think of adding more monitored parameters and widening of family of controlled systems.

Acknowledgements This work has been supported by Project SP2013/168, "Methods of Acquisition and Transmission of Data in Distributed Systems", of the Student Grant System, VŠB - Technical University of Ostrava.

References

1. Nowaková, J., Pokorný, M. *Double Expert System for Monitoring and Re-adaptation of PID Controllers*. Advances in Intelligent and Soft Computing. Ostrava 2013. - accepted
2. Farsi, M., Karam, K.Z., Abdalla, H.H. *Intelligent multi-controller assessment using fuzzy logic*. Fuzzy Sets and Systems. **79**, 25-41, 1996.
3. Nowaková, J., Pokorný, M. *On PID Controller Design Using Knowledge Based Fuzzy System*. Advances in Electrical and Electronic Engineering, **10(1)**, 18-27, 2012. ISSN 1804-3119.
4. Nowaková, J., Pokorný, M., Pieš, M. *Takagi-Sugeno Fuzzy Model in Task of Controllers Design* Advances in Intelligent Systems and Computing, Volume 189 AISC, 2013, pp 391-400. Springer Berlin Heidelberg. ISBN 978-364233017-9.
5. Jacskon, P. *Introduction To Expert Systems (3 ed.)*. Addison Wesley, 1998.
6. Siler, W., Buckley, J.J. *Fuzzy Expert Systems and Fuzzy Reasoning*. Wiley-Interscience, 2004.
7. Jones, Ch.H. *Knowledge Based Systems Methods: A Practitioners' Guide*. Prentice Hall PTR, 1995.
8. Astrom, K. J., Hagglund, T. *PID Controllers: theory, design, and tuning*. USA, 1995. ISBN 1-55617-516-7.
9. Novák, V., Perfilieva, I., Močkoř, J. *Mathematical Principles of Fuzzy Logic*. Kluwer, Boston, 1999.
10. Jager, R. *Fuzzy Logic in Control*. Delft, 1995. ISBN 9090083189.
11. Kilian, Ch. *Modern Control Technology*. Thompson Delmar Learning, 2005. ISBN 1-4018-5806-6.
12. MATLAB - The MathWorks-MATLAB and Simulink for Technical Computing. [cit. 2012-07-10]. <http://www.mathworks.com>.

Remote control of laboratory task using wireless technologies and operating system Android

Vít Otevřel and Zdeněk Slanina

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{vit.otevrel, zdenek.slanina}@vsb.cz

Abstract. This paper is focused to remote control of laboratory task with wireless technologies – Bluetooth and Wifi. As a control part was used device with operating system (OS) Android. Today is OS Android very quick expand in tablets and smart mobile phone. Device with OS Android was used as remote control of laboratory task. As a laboratory task was used model of outdoor blinds. The aim of this task was deepening knowledge about operating system Android and his possibility of communication and wireless technologies Bluetooth and Wifi.

Key words: remote control, laboratory task, outdoor blinds, Android, wireless technologies, Bluetooth, Wifi.

1 Introduction

In recent years, OS Android is very quick expand between public. These devices are mostly smart mobile phones or tablets. Android have support for many communication interfaces so devices with OS Android have usually embedded wireless communication interface – Bluetooth (IEEE 802.15.1) communication module and Wifi (IEEE 802.11) communication module. These devices are more available for public because their price fall down. Android devices have still more and more people so the idea arose about remote control of laboratory task (model of outdoor blinds). This paper continue introduction about both wireless technologies – Bluetooth and Wifi, and next chapter continue to become familiar with the demonstration tasks.

2 Bluetooth

This is a wireless communication technology. Bluetooth is used to connect two or more devices and these devices can communication together. Bluetooth is standard IEEE 802.15.1 and it is belong to the PAN (Personal Area Network). Devices with support Bluetooth communication are usually divided by power (Tabel 1) or data rate (Table 2).[1][2]

Bluetooth works in the ISM¹ frequency band 2400-2480 MHz. During data transfer is used FHSS² method when in one second is performed 1600 retuning between 79 frequency with interval 1 MHz. This mechanism increases resistance to interference on the same frequency.[1][2]

Bluetooth is implemented into many devices – mobile phones, tablets, computer mouses, keyboards, hands-free etc. This technology is used example for quick copy of multimedia content between two mobile phones or tablet and computer. [1][2]

Table 1. Divided devices by power. [1][2]

Class	Maximum allowed power	Approximate range
Class 1	100 mW / 20 dBm	100 m
Class 2	2.5 mW / 4 dBm	10 m
Class 3	1 mW / 0 dBm	1 m

Table 2. Divided devices by data rate. [1][2]

Version	Maximum data rate (Mbps)
Version 1.2	1
Version 2.0 + EDR ³	3
Version 3.0 + HS ⁴	24
Version 4.0	24

3 Wifi

Abbreviation Wifi is marked some standards IEEE 802.11 and these standards describe wireless communication in computer networks. Wifi works similar as Bluetooth in radio frequencies ISM 2.4 GHz. Newest Wifi specification add communication in 5 GHz frequency band. Wifi signal range is bigger than Bluetooth. Wifi devices are divided usually by data rate and version or specification (Table 3). [3]

¹ ISM (Industrial, Scientific and Medical) – frequency band for radio broadcast in industrial, scientific and medical field. This frequency band is free.

² FHSS (Frequency Hopping Spread Spectrum) – method transfer in spread spectrum.

³ EDR (Enhanced Data Rate) – specification Bluetooth 2.0 EDR introduce a new modulation technique $\pi/4$ -DQPSK, for increase data rate.

⁴ HS (High Speed) – high speed data rate conducted through simultaneous Wifi (IEEE 802.11) connection.

Wifi is used usually as replacement classic wire LAN and as medium for connection to internet. Important role in Wifi networks have SSID⁵. This SSID is sent periodically as broadcast ie. potential clients can display available networks on their devices. [3]

Table 3. Divided devices by data rate – most widely used standards. [3]

Wifi specification	Frequency band (GHz)	Maximum data rate (Mbps)
IEEE 802.11b	2.4	11
IEEE 802.11g	2.4	54
IEEE 802.11n	2.4 or 5	450

Ad-hoc mode Type of network when are connected two clients together (peer-to-peer). For identification is used SSID and both clients must be in direct radio range. [3]

Infrastructure mode This is typical Wifi application. In this case is used one or more access points (AP) and AP broadcasts their SSID. It is used for creating local Wifi network or connection to internet. [3]

4 Demonstration task

Model of outdoor blinds was used as demonstration task for remote control. These tasks shown how use Android device for remote control of outdoor blinds over Bluetooth and Wifi technologies.

Model of blinds contains motor which use power supply 230 V/50 Hz. Blinds is controlled by switching phase between inputs on the motor. This procedure allows to change direction of movement up and down. Motor of blinds is automatically turn off when blinds is in end positions (blinds is top or bottom). Model of outdoor blind have not feedback so it is not possible to identify if blinds is in end positions (top or bottom).

4.1 Remote control over Bluetooth

For remote control over Bluetooth was created own main and relay boards. Model of outdoor blinds was connected to relay board which contains SSR⁶ relays. This relay board was connected to main board and Bluetooth module was connected to main board too. Bluetooth module and control application in Android device communication together and main board receives commands. Block diagram is shown on Figure 1. List of used hardware:

⁵ SSID is long 32 ASCII characters. It is used for identification Wifi networks.

⁶ SSR (Solid State Relay) – semiconductor switching element.

- MCU⁷ Atmel ATmega16.
- Bluetooth module EZURiO BTM402 (version Bluetooth 2.0, Class 1) with serial interface.
- Relay board with SSR relays for switching motor of outdoor blinds.

Firmware of main board communication with Bluetooth module over serial interface. For set Bluetooth module was used AT commands. If main board received from control application (in Android device) command over Bluetooth module, then command is decoded and it is switched the relay and movement of motor blinds is changed.

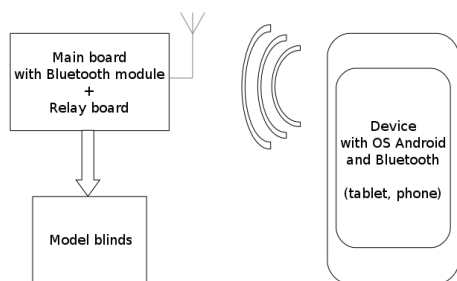


Fig. 1. Block diagram of model blinds with remote control over Bluetooth

4.2 Remote control over Wifi

For remote control over Wifi was used commercial board – LAN module with relay board and Wifi AP. Model of outdoor blinds was connected to relay board. Relay board was connected to LAN module and LAN module was connected to Wifi AP by ethernet cable. Android device with control application was connected to Wifi AP. Over control application in tablet or mobile phone was controlled outdoor blinds. List of used hardware:

- LAN module – contains embedded web server and easy web application. Over this pages are controlled peripherals (relay etc.) on module.
- Module with relay for switching motor blinds.
- Wifi AP/Bridge – dle konfigurace použit jako AP, kdy je k Wifi AP připojeno zařízení s ovládací aplikací přímo, nebo použit jako Wifi Bridge, pro připojení žaluzie do již vytvořené WLAN sítě.

⁷ MCU – micro-controller.

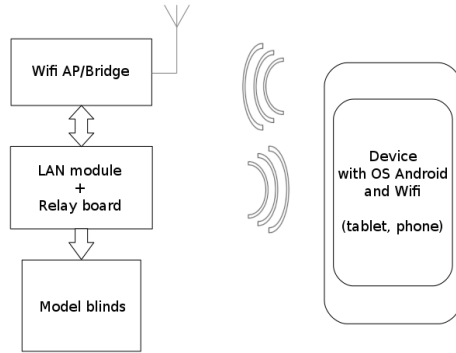


Fig. 2. Block diagram of model blinds with remote control over Wifi

4.3 Control application

Bluetooth Application contains one main screen. This screen contains some buttons for control outdoor blinds. These buttons emulate behaviour of original driver which delivered with blinds. Application contains menu with settings Bluetooth – scan available devices, connect to blinds. GUI of control application is shown on Figure 3 on left.

Wifi Application contains two screens – main screen and settings screen. GUI of control application is shown on Figure 3 on right. Main screen contains three buttons, for direction up and down and stop. From main screen is access to settings over menu. On the settings screen allows set IP address of LAN module and identification numbers of relays used on relay board for change direction of movement blinds.

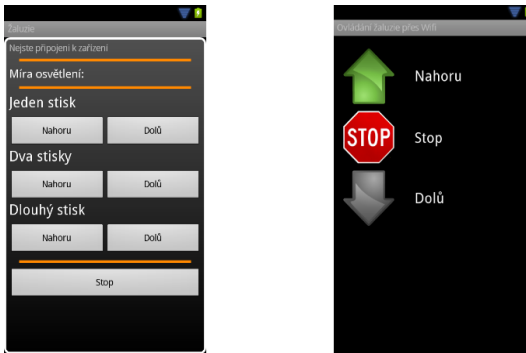


Fig. 3. GUI of control applications for OS Android

5 Conclusion

Bluetooth version was created on own hardware. If board was designed and created so components which were used on boards were choice precisely for this application. So created hardware is small and all important components – MCU, Bluetooth module and relays were embedded on one board.

In other case was used universal LAN module. This is case of control over Wifi. The LAN module have embedded some peripherals such as temperature sensors, A/D converter, relays, some inputs and embedded web server. Over web server is controlled all peripherals on module. For use in our application was necessary watch communication on network between computer and module. The information obtained was used in Android application.

For one concretely application is better designed own hardware because it is smaller and it is designed precisely for this application. If used universal board (example LAN module) it is better for quick measurement or quick launch some application but not for long time usage.

References

1. Bluetooth SIG, <http://www.bluetooth.com/Pages/Basics.aspx>, Basics — Bluetooth Technology Website.
2. Bluetooth SIG, <https://www.bluetooth.org/en-us/specification/adopted-specifications>, Specification — Adopted Documents.
3. Wi-Fi Alliance, <http://www.wi-fi.org/discover-and-learn>, Discover and Learn — Wi-Fi Alliance.
4. Android, <http://developer.android.com/guide/topics/connectivity/bluetooth.html>, Bluetooth — Android Developers.
5. Android, <http://developer.android.com/reference/android/bluetooth/package-summary.html>, android.bluetooth — Android Developers.

Verification methodology for noninvasive continuous measurement of blood pressure

Lukas Peter and Martin Cerny

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{lukas.peter, martin.cerny}@vsb.cz

Abstract. Monitoring of changes in blood pressure is one the most commonly used medical methods. However as standard procedure it can only be done in discontinuous time intervals. Currently continuous measurement of blood pressure is only possible using invasive methods, which makes the measurement impractical for the doctors and uncomfortable for the patients. Physiologically there is connection between electrical and mechanical heart functions and blood pressure value. The article describes the method of measurement and ECG and PPG signal processing so as to obtain information about pulse wave transit related to value of blood pressure. This value is called pulse transit time - PTT. The measurement set up used for simultaneous receiving of all necessary biosignals and measurement software in LabView environment for signal processing of the acquired data are described here. A case study is used for the acquired data and the data are processed by means of regression analysis.

1 Introduction

Continuous measurement of blood pressure is today measured by inserting invasive sensors into the body of a patient, which is uncomfortable for the patient and as all invasive procedures connect with certain risks.

Blood pressure value is connected to electrical and mechanical heart functions, condition of vascular system and transit of biosignals, which are possible to measure without limiting the persons during common daily activities.

At each heart systole the blood from the left ventricle is expelled into aorta and it gives rise to a pressure wave which is further spread around the vascular system. The resulted pressure wave is called a pulse wave and the time in which this pulse wave gets from the heart to other places in vascular system is called pulse transit time - PTT. The PTT value at one place is dependent on the parameters of the blood stream such as elasticity, vascular tissue thickness, vessel diameter and blood pressure value.

There are more ways to detect PTT. One of them is to measure the time delay between R-wave from ECG recording and the initial point of another pulse wave.

These parameters can easily be acquired from commonly measured biosignals (ECG, photoplethysmography - PPG).

The sensors for measuring biosignals necessary for PTT determination are small in size, they are easy to firmly attach to the patient's body and do not bother the patient much.

2 The current state analysis

The method of CNIBP measurement based on PTT is described in many scientific articles, for example [1], [2]. There exists several ways to assess blood pressure value that are based on PTT. The blood pressure evaluation using PTT is dependent on the method of taking blood pressure.

The algorithms used in the articles describing CNIBP measurement with which it is possible to determine PTT with a certain accuracy[1], [2], are always dependent on constants describing hemodynamic properties of the cardiovascular system.

$$\Delta P = \frac{2}{T_y} + PTT \quad (1)$$

$$\begin{aligned} P &= \frac{1}{0,7} \left(\frac{1}{2} \rho \frac{d^2}{PTT^2} + \rho gh \right) \quad (2) \\ &= \frac{A}{PTT^2} + B \end{aligned}$$

Where T_y in the first formula [2] and A, B in the second formula [1] is characteristics of the cardiovascular system.

But these constants are not clearly described and their determination is also not very clear. The hemodynamic properties of the cardiovascular system will differ with each monitored person and therefore it will be very difficult to determine their general evaluation.

An effective method for measuring blood pressure based on PTT and an algorithm for evaluating for blood pressure value would diagnostically contribute to improving medical care and at the same time it would help to anticipate many physiologically dangerous states, which are caused by value of blood pressure [3].

3 The method of measurement

PTT is acquired as the time delay between R-wave from ECG recording and the initial point of another pulse wave. It is therefore necessary to take biosignals from ECG and PPG and PTT correlation was performed to the values of invasive blood

pressure. The measurement set up was constructed in order to measure all these signals.

During the test of measurement set up standard single use AgCl adhesive electrodes for ECG measuring together with a sensor for PPG signal were attached to the monitor person. Taking the blood pressure was performed under supervision of a specialist M.D. because the patient was catheterised during the measurement.

All the data from measurement are saved in dedicated software and ready for consequent processing and evaluation.

4 Measuring device design

The whole measurement set up could be perceived as two individual parts.

The hardware part, which is made up of individual sensors and probes taking electrical and nonelectrical impulses from the body of the patient, amplifier of biological signals, hardware for measuring of pulse wave and AD converter.

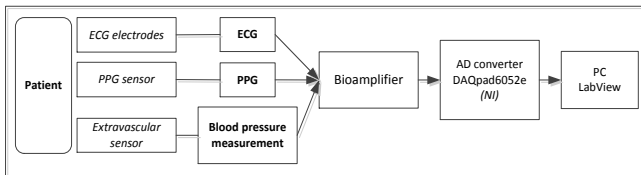


Fig. 1. Measuring chain

4.1 PPG

A specific part of measurement equipment was designed for taking the photoplethysmography signal from an ear lobe. PPG wave was recorded using the probe working on the principle of transmission.

The equipment was constructed for universal taking of PPG wave at various parts of the body using a sensor with a compatible connector CAN9.

The whole equipment is powered by two 9-volt batteries by which symmetrical power supply of the equipment is insured and it is not necessary to galvanically separate the power supply from the patient's circuit.

4.2 BP

Blood pressure is measured by an invasive method. An extra vascular sensor is used. This sensor works on a principle of measurement on a Wheatson resistance bridge and it is always connected directly into the monitoring equipment. It was necessary to create hardware interface for measuring the signals at resistance bridge and its consequent connection to AD convertor. Extravascular sensor DPT-6100 was used.

Because the measurement at the resistance bridge is done in relation to two terminals, it was needed to bring the signal to a value related to the earth terminal. In this step amplifier INA126 was used. This amplifier is powered by symmetric voltage. Signal from the resistor bridge is brought to input terminals of the amplifier and the output is amplified difference of the entry signals, i.e. signal related to earth terminals.

It is necessary to insure stable and accurate power supply of the resistance bridge in order to measure the blood pressure precisely. This power supply is provided by a stabiliser mode from stabilizing diode TL431, which is characteristic by its temperature stability, by which it insures stability of power supply and at the same time it is possible to alter the voltage to a small extent.

After connecting the sensor it was necessary to calibrate it on an acquired pressure value. These calibration was achieved using a laboratory calibrating device when the acquired pressure value was set and then the voltage coming from chain measuring equipment for measuring blood pressure was subtracted.

5 Software

All the measurement software was developed in LabView environment. The software is divided into two parts: a part for reception and saving of data and a part for data processing and evaluation.

The whole software is primarily designated for saving measured data for consequent processing but nonetheless it is also equipped by a visualisation part, which contains simple filtering of the received signals, which are consequently displayed on a monitor.

The signal from a pressure sensor is processed entirely separately. For blood pressure measurement it is necessary to use zero adjust. In the moment when the sensor is connected to the applied catheter it is necessary to subtract the immediate voltage value. This value is generated on the sensor before opening of a three-way faucet, by which the sensor is equipped, and it corresponds to absolute pressure, i.e. offset sensor.

The signal was filtered by Low pass filters, the only signal not filtered by an low pass filter is a signal from the pressure sensor, which is filtered by a simple filter for medium values.

Low pass filter 30 Hz was used for filtration of ECG signal. The filter 15 Hz was used for filtration of PPG signal. The values of the filters were set so as to avoid filtering out the needed frequencies, i.e. 8-22 Hz for QRS and 0,5-15 Hz for PPG. This part of the programme is created mainly for detection of significant features of individual signals, i.e. the R-wave of ECG signal, the peaks of PPG signal and the peaks of pressure. Wavelet transformation was used to detect the R-wave of ECG signal. Filter Daubechies 4 was chosen as the most suitable decomposition filter used for WT in the detector.

6 Description of the experiment

Signals from ECG, PPG and a pressure wave were sensed in 8-minute intervals. Blood pressure was measured in femoral artery, where patient was catheterized. At the end of the catheter extravascular sensor was attached. 4-lead ECG was measured from the chest. PPG pulse wave was measured using the sensor at the left ear lobe.

The sampling frequency was set to 1 kHz during the measurement.

The data were saved into .txt format. The program is also ready to save the data in a binary code.

After the consequent analysis of the measured data the detection of peaks in the whole measured interval was done for all the measured signals. In order to statistically evaluate the measured data, when the values of PTT and blood pressure values are available and the dependency of the PTT and changing blood pressure value is sought, it is suitable to use a method of regression analysis according to the information search above.

A regression analysis was done regarding all the possible ways of PTT determination as a time delay between the R peak of ECG and either the start or the peak of the pulse wave. Furthermore analysis for particular parts of the period between the start and the peak of the pulse wave was performed. Results of those analyses showed that the most accurate value of determination of PTT from the pulse wave is neither the start nor the peak but a 30% value between these points.

During the analysis several windows for moving average of different length from the smallest 3-segment to the largest 15-segment window. The segment here is the period from one R-wave to another.

7 Results of the experiment

The regression analysis for all windows showed that there is a linear dependency between PTT and blood pressure. With a rising window of a moving average the dependency between PTT and blood pressure was also rising.

Table 1. Results of the regression analysis

MA [segment]	Corelation coefi- cient	R-square [%]
3	0,41	16,43
5	0,49	24,91
11	0,65	42,15
15	0,68	46,63
60	0,82	68,17

It is obvious that if the window of moving average was larger it would be possible to reach a high degree of accuracy of blood pressure determination. Nevertheless

after a consultation with doctors from a cardiology department of the University Hospital of Ostrava it was clear that the diagnostically suitable time period is 10-20s maximum delay, which corresponds in this case to the 11- and 15-segment window.

8 Conclusion

All the data were acquired only during the examination of one patient, and thus it is not possible to claim that the results of the statistical analysis are generally valid. Nonetheless the complex hardware and software solution created in this work for laboratory analysis of NIBP is ready for further tests leading to confirmation and development of measurement method of NIBP.

In the next step a clinical study will be carried out, result of which will be a sufficient database of measurements. Thanks to these acquired data a further development of the measurement method of NIBP using new methods of modern mathematics will be possible.

The outcome of the further work will be new algorithms precisising NIBP measurement.

References

1. PARRY, F, G DUMOT, C RIES, C MOTT, M ANSETMINO. Continuous noninvasive blood pressure measurement by pulse transit time. EMBC 2004: 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society : conference proceedings : linkages for innovation in biomedicine : 1-5 September, 2004, San Francisco, California. Piscataway, N.J.: IEEE, c2004. ISSN 0-7803-8439-3.
2. DINGILI, Li, Yi PAN, Hang CHEN, Ye SHUMING, Yan HONG. The Establishment of a Non-Invasive Continuous Blood Pressure Measure System Based on Pulse Transit Time. *The 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE 2008): May 16-18, 2008, Shanghai, China [online]*. Piscataway, N.J.: IEEE Xplore, 2008 [cit. 2012-07-17]. ISSN 978-1-4244-1747-6.
3. MA, T, Y.T. ZHANG. A Correlation Study on the Variabilities in Pulse Transit Time, Blood Pressure, and Heart Rate Recorded Simultaneously from Healthy Subjects. *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. [s. l: s. n.]. ISSN 0-7803-8741-4.
4. AUGUSTYNEK, M, PENHAKER, M. Non Invasive Measurement and Visualizations of Blood Pressure. In: *Elektronika ir elektrotechnika / Kauno technologijos universitetas, Lietuvos Mokslų akademija [ir] Vilniaus technikos universitetas*. ISBN 1392-1215

Testing gapless measurement of PQA according to IEC 62586-2

Petr Pětvaldský

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.petvaldsky@vsb.cz

Abstract. This paper describes a tester for automated testing of power quality analyzers, focuses on evaluation of correct gapless and non-overlapping measurement according to standard IEC 62586-2. A test is used in order to verify the accuracy of the basic time window for evaluating RMS voltage values (10/12 cycle), and to evaluate whether during the measurements RMS samples (gapless) are not lost and mutual overlaps of individual cycles are not produced. The tester is based on principles of virtual instrumentation. The testing hardware uses PXI system architecture and the software is made in NI LabVIEW programming environment.

1 Introduction

Due to large expansion of renewable energy resources in distribution grid, occurrence of disturbances in the power system has been increasing during few last years. It leads to degradation of power quality in power system. Therefore, long-term monitoring and evaluation of power quality is becoming increasingly important not only for electricity distributors but also for European regulators and customers.

With regard to these facts, systems for measurement and analysis of power quality are being increasingly used in different places of power system. Although the requirements for evaluation of power quality are described in IEC 61000-4-30 [2], testing equipment for power quality analyzer testing is not currently available. Last year (2012) a new international standard IEC 62586-2 [1] was published which describes new demands for testing power quality analyzers.

The Faculty of Electrical Engineering and Computer Science, VSB-Technical University of Ostrava is currently developing fully automated equipment that would be able to test devices which measure the power quality. This paper is focused on the description of testing correct gapless and non-overlapping measurement.

2 Gapless (non-overlapping) measurement

PQ analyzers compute RMS values from the basic measurement interval of 10/12 cycles (10 cycles at the signal frequency 50 Hz and 12 cycles for the frequency of 60 Hz) which is in both cases 200 ms time period. These values are then aggregated to continue the calculations for the following three periods:

- 150/180-cycles interval,
- 10-minute interval,
- 2-hour interval.

Individual intervals should be connected to each other (gapless) and should not overlap, apart from exceptions in the time synchronization every 10 minutes. Last 10/12 cycles interval in 10-minute intervals usually overlaps over the next 10/12 cycles interval, this interval is then included in the aggregation of the previous 10-minute interval.

The exact timing of each interval and the synchronization is not an easy task. Thus there is a suspicion that some power quality analyzers manufacturers can simplify the measuring algorithm, so that an incorrect implementation is not visible at first sight. An example of incorrect implementation of 10/12 cycles measurements can be using a moving window of exact length 200 ms (or even greater than 200 ms) although it is required that the RMS values should evaluate for the 10/12 cycles (a frequency deviation from the 50/60 Hz cause inaccuracies).

3 Test hardware description

In terms of hardware the test system consists of a control computer which is running the test software, which will be described in the next chapter.

The control computer is present here as a PC based device in PXI chassis (PCI eX-tensions for Instrumentation). PXI system, which is a part of the tester is marked NI PXI 1033. It is a standard PXI chassis, which can be fitted with up to five extension modules.

Only one expansion module is used and it is a multifunction NI PXI 6733 card, which allows generating analog output signals with sampling frequency up to 1 MS/s. The card is equipped with D/A converter with 16-bit resolution.

With this multifunction card signals are generated for particular tests, as prescribed by the standard. This card, however, allows generating signals only in the range of -10 V to +10 V. Therefore the generated signal is amplified by power amplifiers to the level of 65 V, which corresponds to the declared input voltage (U_{din}) range of a tested power quality analyzer. Thus, the amplified signals for all the three phases can then be routed to the measuring inputs of the power quality analyzer. For purposes of this particular test part usage of only one phase was sufficient.

In Fig. 3 below, please see the block diagram and photography of the real test system, which is stored in mobile 19" rack.

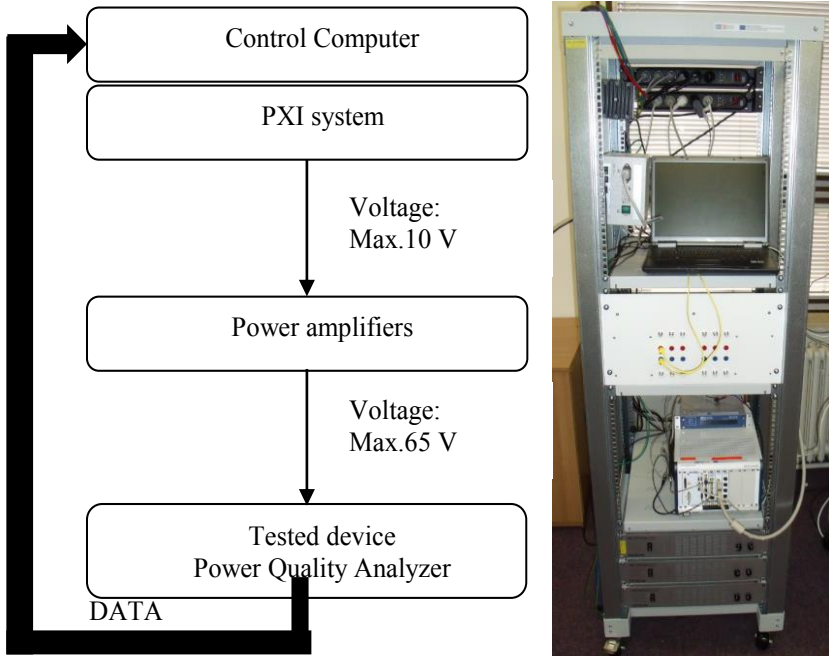


Fig. 3. Block diagram and photography of test system

4 Test description

For testing the correct evaluation of the RMS amplitude, the device under test must be tested with fluctuating signal of fundamental frequency with the following parameters:

- Sinusoidal modulation,
- Basic harmonic component amplitude: 100 % U_{din} , (V1)
- Basic harmonic component frequency: 50 Hz, (f1)
- Modulation depth +/- 10 %, (A_m)
- Modulation frequency: 2,3 Hz. (f_m)
- Phases: Not required. (φ_1, φ_m).

Signal can be mathematically expressed as follows:

$$s_{RMS}(t) = V_1 \sqrt{2} \cos(2\pi f_1 t + \varphi_1) (1 + A_m \cos(2\pi f_m t + \varphi_m)). \quad (1)$$

With the above parameters for the voltage U_{din} 230 V signal would be as shown in Fig. 4.

In this ideal case (theoretical simulation) the fluctuation frequency is exactly 2.3 Hz. It is now relatively easy to detect gaps or overlaps during measurement using FFT transformation. The spectrum is calculated from the relative (measured RMS value to U_{din}) RMS values (array of 100 elements).

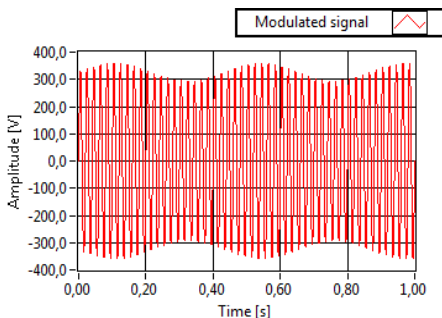


Fig. 4. Modulated (testing) signal

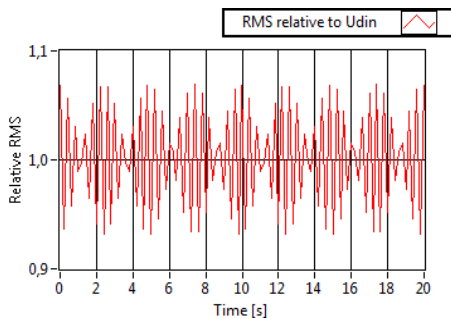


Fig. 5. RMS values relative to Udin (simulation)

If only one missing sample of RMS value in the analyzed array exists, spectral analysis reflects with spectrum leakage effect. It can be seen in the graph in Fig. 6. Blue color shows the spectrum that corresponds to the set of values, in which there are no missing values. Red color shows the spectrum of array that is missing one sample RMS value. The same disproportion in measured values as missing sample causes the gaps or overlaps.

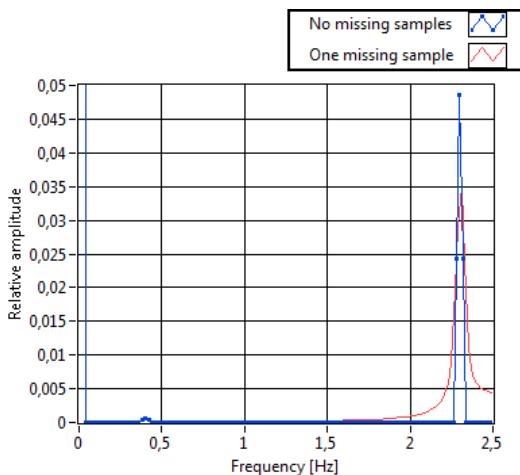


Fig. 6. FFT analysis of RMS variations (theoretical simulation)

The test should not take longer than 10 minutes, as this could lead to a possible overlap due to 10-minute aggregation interval. This, in this case, will be correct.

Voltages U_{din} (Declared input voltage) with an ideal signal / noise ratio should be applied to the test. Manufacturer of test equipment itself should specify the optimum value of this voltage.

The tested device should provide in interval of every 10/12 periods measured value of RMS voltage with a time stamp with the history of the samples with a length of at least 100.

The test device must store the data in a log file, or be able to read the measurement data from the communications port in order to subsequently process the measured data for testing purposes.

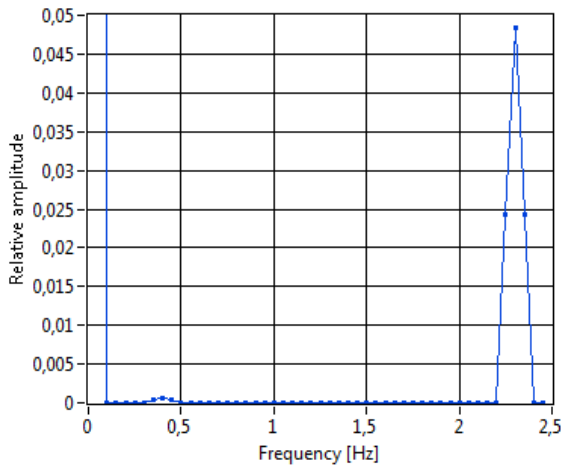


Fig. 8. FFT analysis of RMS variations (PQA ENA330)

Fig. 8 shows the spectral analysis of values measured by ENA330 analyzer. From the spectral analysis can be concluded that the analyzer meets the terms of the requirement for gapless measurement because there is no visible spectral leakage. But to obtain probative value, it is necessary to perform additional calculations.

5 Evaluation of results

Measured sequence of RMS values from 10/12 cycles period of signal are $U_{rms}(0) \dots U_{rms}(99)$. From this sequence it is possible to calculate the following values:

$$A(N) = \left\| \frac{1}{50\sqrt{2}} \sum_{n=0}^{99} U_{RMS}(n) e^{j2\pi \frac{n}{N} k} \right\|, k = 45,46,47 \dots (2)$$

Note: The double line denotes modulus of a complex number.

Where N size of an array with measured RMS values, k is the position of the FFT bin corresponding to the modulation frequency, and $A(k)$ is the amplitude of the bin.

As an indicator of missing samples, or overlaps, the following formula can be used:

$$Q = \sqrt{\frac{A(k)^2}{A(k-1)^2 + A(k+1)^2}} \dots (3)$$

In this case, the modulation frequency of 2.3 Hz and the field size, determined for FFT analysis, 100 RMS values, the bin corresponding to modulation frequency is $k=46$, assuming that the DC component is bin with index 0.

$$Q = \sqrt{\frac{A(46)^2}{A(45)^2 + A(47)^2}} \quad (4)$$

According to IEC 62586-2 [1] following requirements should be met:

- 1) $Q > 20$
- 2) $4.5 \% < A(46)/U_{din} < 5.5 \%$
- 3) $\text{Time}(U(99)) - \text{Time}(U(0)) = 20 \text{ s } \pm 6 \text{ ms}$

Analyzing the data measured by the analyzer ENA330 when testing under the above conditions, we came to the following values:

Table 1. Test result values.

A(45)	A(46)	A(47)	A(46)/U _{din}	Q	t(U(99))-t(U(0))
0,0045	3,16841	0,0073	4,8424 %	369,4703	19,997

6 Conclusion

Analyzing the results of the test, it was found that the measurement algorithm of tested analyzer ENA330 is processed appropriately according to the requirements of IEC 62586-2. Another analyzer (LEM Memobox 300) was also tested, but was found that this device does not export values with 200 ms interval, the lowest interval available for export values was measured 5 minutes. This is insufficient for analysis, so the unit had to be excluded from testing. The test described in this article is part of a comprehensive tester, it's still in development. Result will be a tester for automated testing of power quality analyzers.

References

1. IEC 62586-2 Ed.1. Power quality measurement in power supply systems: Part 2: Functional tests and uncertainty requirements. Geneva: International Electrotechnical Commission, (2012).
2. IEC 61000-4-30 Ed.2.0 Electromagnetic compatibility (EMC): Part 4-30: Testing and measurement techniques – Power quality measurement methods. Geneva: International Electrotechnical Commission, (2008). ISBN 2-8318-1002-0.
3. IEC 61000-4-7:2002/A1:2008 Electromagnetic compatibility (EMC): Testing and measurement techniques - General guide on harmonics and interharmonics measurements and instrumentation, for power supply systems and equipment connected thereto. Geneva: International Electrotechnical Commission, (2008).
4. IEC 61000-4-15 Ed.2.0. Electromagnetic compatibility (EMC): Part 4-15: Testing and measurement techniques – Flickermeter – Functional design specifications. Geneva: International Electrotechnical Commission, (2010). ISBN 978-2-88912-076-5

System for Automatic Loading under Operating System QNX

Roman Slabý and Martin Stankuš

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{roman.slaby, martin.stankus}@vsb.cz

Abstract. Presented paper describes the possibilities of drivers loading in QNX real-time operating system. System for automatic driver loading was developed, because communication fieldbus often requires setting of specific parameters related to bus communication and drivers under QNX are not loaded automatically but are loaded by shell script or console command. System for driver loading is based on database implemented using SQLite technology and information about cards connected to the bus stored in the PCI card configuration space.

Key words: QNX, Driver, Photon SQLite, PCI Card.

1 Introduction

QNX operating system is multi-platform, real-time operating system that uses microkernel architecture and is developed by QNX Software Systems. QNX is a Unix-like operation system compliant with POSIX specification [1].

Graphical configuration system is designed to manage software support for industrial communication adapters in the QNX operating system. Individual industrial communication networks differ in needed software support and methods of communication. Graphical configuration system unites these different requirements for software support and configuration.

Graphical configuration system can be divided into several functional blocks. The core of the system is a database that stores all configuration data. This database is implemented using SQLite technology. The data from the database are displayed and modified in the graphical user interface. Each industrial communication adapter is provided with dedicated space in the graphical user interface. Script management module performs operations with individual scripts, such as calling of scripts with appropriate parameters. These operations are based on the data from the database. Initial configuration is loaded into database by scripts specific for every type of industrial communication adapter [1], [2].

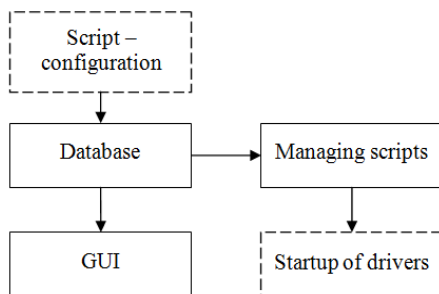


Fig. 1. Block diagram of the automatic driver loading.

2 Identification of connected cards

Graphical configuration system allows setting of configuration parameters for supported cards and is able to load the drivers with appropriate parameters for these cards. Identification of cards begins with generation of data file containing output of "enum-pci" command. Information describing individual cards are separated from the data file and compared with the information stored in the database. Sequence of driver loading is shown in Figure 2.

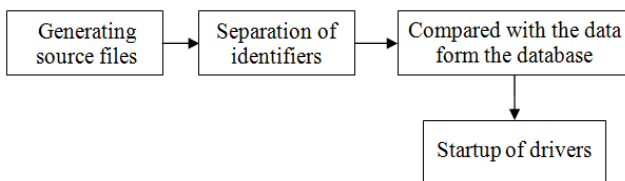


Fig. 2. Identification of the PCI card.

Separated data containing the identifiers of all available cards are compared with the data stored in a device database. Identifiers are information needed to uniquely identify every attached card. These identifiers include Vendor ID, Devices ID, Subsystem Vendor ID and Subsystem ID. These values are stored in the PCI card configuration space. Vendor is unique 16 bit number used to identify manufacturer of the card. This unique number is assigned for a fee to every PCI device manufacturer by PCI consortium, which involves more than 800 companies. The Board of Directors of this consortium is composed of representatives of the most important technological companies, such as Intel, Microsoft, IBM, AMD, Agilent Technologies and more. Device ID is assigned by device manufacturer. This 16 bit value is unique within particular Vendor ID scope and is stored in Device ID register. There is possibility to use Subsystem Vendor ID and Subsystem ID to further specify card's manufacturer and model of the card [3].

Unlike databases build on a client-server model, where the database server is running as a dedicated process, SQLite is only a small library that is linked with the application. Each database is stored as a separate .dbm (Database Manager) file, where the data are stored in equally sized blocks using a simple primary key. Database for the configuration system contains two tables. These tables are named DEVICES_INFO and DRIVER. Table DRIVER contains information about executed driver and its parameters. Table DEVICES_INFO contains list of supported cards (devices) and data elements to uniquely identify the card. Besides this it contains STATUS flag determining whether a given card is present or not [4], [5].

Table 1. Contents of the table DEVICES_INFO.

DEV_ID	VENDOR	DEVICE	SUBVENDOR	SUBSYSTEM	DEC_CH1	DEC_CH2	DEC_CH3	DEC_CH4	STATUS
1	3756	8	3756	8	2	0	2	0	1
2	4277	36912	5280	42	1	0	0	0	0
...	28	1	28	1	1	0	0	0	0

Driver loading is realized by script management block. This block continuously monitors database transactions. Actual driver loading or unloading is implemented by shell scripts. Calling of these scripts is based on database monitoring by script manager block. Script filenames, paths to scripts and scripts parameters are stored in a database table named DRIVER.

Table 2. Contents of the table DRIVER.

DRV_ID	COMMAND	PARAMETERS	LOAD
1	"/sbin/runme.sh"	"-x 500 -t 10"	1
2	"/sbin/runme1.sh"	"-x46 -t 12"	0
...	"/sbin/runme.sh"	"-x 25 -t 150"	0

Database content is read after start of the system. This is followed by loading of the driver for every card, which is described by a valid entry in the database and is also physically connected to the computer. Database content can be dynamically modified by user changes through GUI. From a technical view the application is notified by calling of callback function. This callback function identifies the record in the database which is related to change and performs

update of this record. Termination and reloading of driver with new parameters is performed as well.

Drivers are unloaded before termination of system operation. Individual scripts for driver termination are called with appropriate parameters.

3 Graphical user interface

Parameters for every driver can be changed in the Graphical User Interface created in Photon. Photon provides a flexible, easy-to-use environment for you to interact with your computer. Like QNX itself, Photon is built around a small microkernel. This modular architecture makes Photon fast, flexible, and inherently capable of network-distributed computing. It's designed to fit in embedded systems.

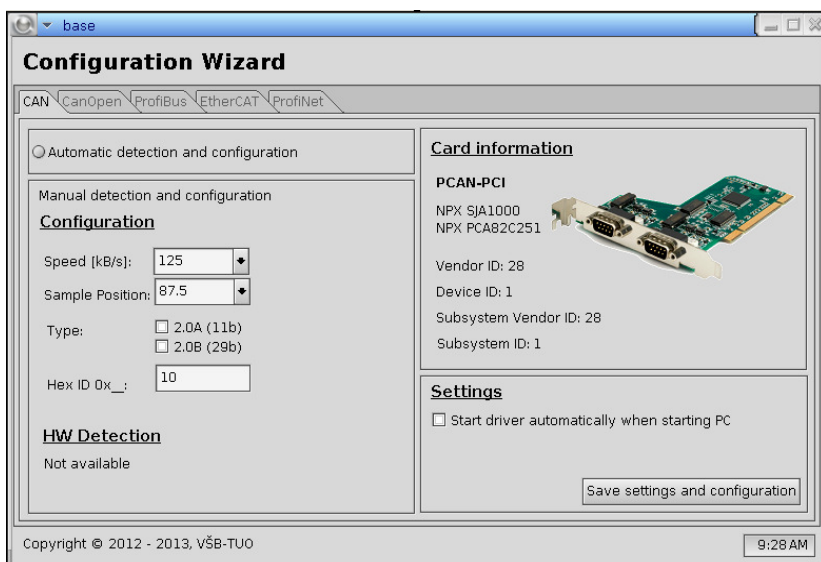


Fig. 3. Graphical user interface for configuration of the CAN driver

4 Conclusion

Automatic driver loading is matter of course for users of the Windows operating system. However, this is not true for users of QNX operating system. It is necessary to load the driver by shell script or console command under this operating system. The goal of described system is to clarify and facilitate initial setup of drivers including their loading. Our graphical configuration system currently supports only limited number of communication cards but can be extended.

This extension can be especially useful for systems like SCADA. Communication via fieldbus networks often require setting of specific parameters related to bus communication. The disadvantage of our system is the need to adjust the graphical interface whenever the set of supported cards is changed. This could be solve by dynamic modification of graphical user interface with layout of interface describes by additional data.

Acknowledgements This work has been supported by Project SP2013/168, "Methods of Acquisition and Transmission of Data in Distributed Systems", of the Student Grant System, VŠB - Technical University of Ostrava.

References

1. KOLNICK, Frank. *The QNX 4 real-time operating system: programming with messages in a distributed environment. new ed.* Willowdale, Ont: Basis Computer Systems, 1998. ISBN 09-219-6001-8.
2. G KRTEN, By Rob. *Getting started with QNX Neutrino 2: a guide for realtime programmers.* Kanata, Ont: PARSE Software Devices, 2001. ISBN 09-682-5011-4.
3. ABBOTT, Doug. *PCI bus demystified.* Eagle Rock, Va.: LLH Technology Pub., viii, 198 p, 2000. ISBN 18-787-0754-X.
4. KREIBICH, Jay A. *Using SQLite. 1st ed.* Sebastopol, CA: O'Reilly, 2010. ISBN 05-965-2118-9.
5. QWENS, Michael. *The definitive guide to SQLite.* New ed. Berkeley, CA: Apress, 2006. ISBN 15-905-9673-0.

The SSVEP Detection Using the Canonical Correlation Analysis and the Continuous Wavelet Transform

Maros Smondrk, Michal Vozda, Jana Krohova and Barbora Hrvolova

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{maros.smondrk, michal.vozda, jana.krohova, barbora.hrvolova}@vsb.cz

Abstract. The steady state visual evoked potential (SSVEP) is one of the most discussed method for the brain computer interface (BCI). In this paper, performance of the online SSVEP detection methods was compared. Two nonparametric methods were used: the canonical correlation analysis (CCA) and the continuous wavelet transform (CWT). The success of the SSVEP detection and the time consumption of these methods were compared.

1 Introduction

The brain–computer interface (BCI) represents communication and control interface between the brain and an external device. In particular, it provides transformation of user’s intent conveyed by brain (electroencephalogram (EEG)) signals to set of machine (device) commands. The main feature of the BCI is that the generation of brain signals does not depend on neuromuscular activity. [1]

Significant progress has been made in researching the BCI technologies in the last decade. The applications controlled by these interfaces have large potential as assistive technology for users with severe disabilities (e.g. amyotrophic lateral sclerosis or brainstem stroke), users with neuromuscular impairment etc. [1]-[3]

The most utilized electrical activities of the brain for BCI are beta and mu rhythms, P300 evoked potentials, visual evoked potential, slow cortical potentials and steady-state visual evoked potentials (SSVEP). [4]

In this paper, we focused on the SSVEP. These brain signals are natural response of the retina to a repetitive visual stimulation with the frequency from 3.5 Hz to 75 Hz. Several researches state that the best sensitivity of the retina is around 15 Hz. SSVEP has the same fundamental frequency as the visual stimulus (flickering light source). It also consists of the first, the second and higher harmonics. [5]-[7]

There are number of SSVEP processing methods for detection and extraction of the fundamental flickering frequency. In this study, performance of two methods for SSVEP detection is compared: the canonical correlation analysis (CCA) and the continuous wavelet transform (CWT).

2 Experimental Setup

Ten volunteers (8 males and 2 females) at the age of 20 to 30 participated in the experiment. All of them have normal or corrected-to-normal vision. They were seated about 100 cm in front of the professional stroboscope (which is used during the EEG measurement) in a comfortable office chair in a slightly darkened and quiet room. All of the volunteers had never used the SSVEP-based BCI. Each subject participated only in one measuring run. The stimulation frequencies were 17, 19, 21, 23 and 25 Hz. The measurement run consists of five stages. In each stage, one stimulation frequency was used. Duration of the stage was 60 s and between each stage a volunteer took a break for 30 s.

Many studies have shown that the best locations to detect and acquire SSVEPs are the occipital areas of the brain [8]-[10]. The EEG signals were measured from two bipolar channels O1-Pz and O2-Pz placed in 10/20 international system (Fig. 1) of electrodes with the USB biosignal amplifier – g.USBamp from g.tec. The sampling frequency was set up to 256 Hz, with a notch filter at 50 Hz. The impedances of the passive electrodes were kept below 5 k Ω .

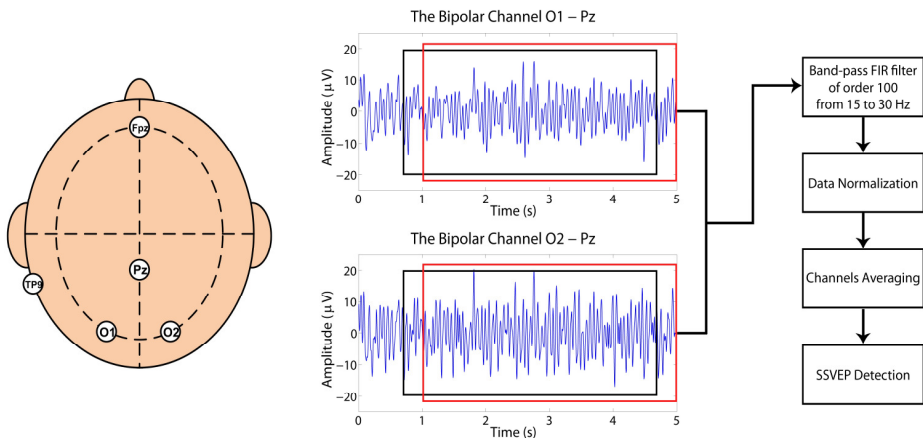


Fig. 1. The electrode placement (O1-Pz and O2-Pz are two bipolar channels, Fpz is a ground electrode and TP9 is a reference electrode), the example of the measured signals from the bipolar channels and the signal processing block diagram. The red and black boxes represent the time window of 4 seconds moving each 250 milliseconds.

The 1024 samples of EEG signal were acquired and processed in every loop but only 64 samples of new data was added. Data was then filtered using a band-pass finite impulse response (FIR) filter of order 100 from 15 to 30 Hz. The filter was designed using the Matlab toolbox. After filtering the data was normalized and two bipolar channels were averaged to one channel for all detection methods. Subsequently, data was processed by detection methods which are discussed in the next session.

3 Methods

3.1 The Canonical Correlation Analysis

Canonical correlation analysis is a multidimensional statistical method measuring the linear relationship between two sets of variables \mathbf{x} and \mathbf{y} . CCA determines a set of canonical variates, orthogonal linear combinations of the variables within each set that best explain the variability both within and between sets. In our case a set of variables \mathbf{x} is represented by EEG signal recorded from two bipolar channels O1-Pz and O2-Pz placed in 10/20 international system of electrodes. The set of variables \mathbf{y} is created from K harmonics of reference signal.

$$\mathbf{y} = \left(\sin\left(2\pi f \frac{n}{S}\right) \quad \cos\left(2\pi f \frac{n}{S}\right) \quad \cdots \quad \sin\left(2\pi Kf \frac{n}{S}\right) \quad \cos\left(2\pi Kf \frac{n}{S}\right) \right)^T, \tag{1}$$

where f is wanted frequency of stimulus in EEG, K is the number of harmonics that are taken into. The n is the number of the sampling points and S is the sampling rate. [9], [11], [12]

In practice, during the calculation of canonical correlation the covariance matrix from sets of variables \mathbf{x} and \mathbf{y} is created.

$$\mathbf{C} = E \left[\begin{pmatrix} \mathbf{x} - E[\mathbf{x}] \\ \mathbf{y} - E[\mathbf{y}] \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x} - E[\mathbf{x}] \\ \mathbf{y} - E[\mathbf{y}] \end{pmatrix}^T \right] = \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix}, \tag{2}$$

where sets of variables \mathbf{x} and \mathbf{y} are subtracted by their average value. \mathbf{C} is a block matrix where \mathbf{C}_{xx} and \mathbf{C}_{yy} are the within-sets of covariance matrix of \mathbf{x} and \mathbf{y} variables. $\mathbf{C}_{xy} = \mathbf{C}_{yx}^T$ is the between-sets covariance matrix. The canonical correlations can be found by solving the eigenvalue equation, [13]-[15]:

$$\mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} \mathbf{C}_{yy}^{-1} \mathbf{C}_{yx} \mathbf{w}_x = \lambda \mathbf{w}_x, \tag{3}$$

where λ is the eigenvalue and it is proportional to the square of the canonical correlation $\rho = \sqrt{Re(\lambda)}$. The vector \mathbf{w}_x is a normalized canonical correlated basis vector. Only maximal eigenvalue is important for SSVEP detection and we found it by the power iterative method.

The CCA method was computed in range from 15 Hz to 30 Hz with 0.1 Hz step for the first bipolar signal from O1-Pz channel. The matrix of reference signals was created for this algorithm. The length of matrix of reference signals is the same as length of segment of measured signal. Reference signals were generated for three harmonics of basics sinus and cosines signal. The output of CCA is a set of canonical correlation coefficients. These coefficients represent correlation between the input signal and the created harmonics signals of particular frequencies. Using these coefficients, the frequency with the maximum correlation coefficient was considered as the dominant frequency of the signal.

3.3 The Continuous Wavelet Transform

In 2010, the CWT was used for detection of SSVEP for the first time [16]. In this article, comparison with the CCA method used in the BCI for SSVEP detection is made. The time-frequency description of the signal is enabled by the CWT, and good time and frequency localization of events are offered, too.

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt. \quad (4)$$

The transformed signal $W(a, b)$ is a function of the translation parameter b and the scale parameter a . The mother wavelet is denoted by ψ , the * indicates that the complex conjugate is used in case of a complex wavelet. The signal energy is normalized at every scale by dividing the wavelet coefficients by $1/\sqrt{a}$. This ensures that the wavelets have the same energy at every scale.

In SSVEP detection, the best results are obtained by complex Morlet wavelet [16], which is determined as a harmonic function modulated by Gaussian window. Morlet wavelet is symmetric, complex and it is not an orthogonal wavelet. Complex Morlet Wavelet is determined by this formula:

$$\psi(t) = \pi^{-\frac{1}{4}} \left(e^{ia_0 t} - e^{-\frac{1}{2}a_0^2} \right) e^{-\frac{1}{2}t^2} \rightarrow \psi_{ab}(t) = \frac{1}{a} \pi^{-\frac{1}{4}} e^{-ia_0 \frac{t-b}{a}} e^{-\frac{1}{2} \left(\frac{t-b}{a} \right)^2}. \quad (5)$$

To implement the discretized version of the CWT, we assume that the input sequence is a length N of vector \mathbf{x} . The discrete version of convolution is described:

$$W[a, b] = \sum_{n=0}^{N-1} x[n] \psi_{ab}^*[N-n]. \quad (6)$$

To obtain the CWT, the convolution has to be computed for each value of the shift parameter b , and this process has to be repeated for each scale a . Values of scale a determine the degree to which the wavelet is compressed or stretched.

A scalogram of measured signal is determined by square of wavelet coefficients. At the edges of the interval, the border distortion is caused by the finite length of investigated signal. One hundred coefficients of each side of the interval are removed to eliminate the border distortion effect. Resulting 1D spectrum is obtained by averaging individual coefficients for designated frequency bands. Using the 1D spectrum, the frequency with the maximum spectral energy was considered as the dominant frequency of the signal.

CWT method was computed with scale factor from 1 to 29 with step 1. This scale determines frequency range from 13.24 Hz to 384 Hz with different steps. The frequency range and frequency bands depend on the kind of wavelet. The complex Morlet wavelet with bandwidth parameter 1 and center frequency 1.5 was used for detection of SSVEP.

4 Results and discussion

In this paper, we compared the success and time consumption of detection of two methods. The time consumption of the CCA method is 52.6 ± 0.7 ms and CWT method is 114.2 ± 2.8 ms. Computations were performed in Matlab 2011a environment on a PC with i7core 2x2.8 GHz, 4 GB RAM and 32bit OS Windows 7. Taking into account that we processed 64 new data samples in each loop with sampling frequency 256 Hz, the maximal computation time of any detection method has to be less than 250 ms in real-time application. The results show that the computation cost is negligible. The main goal was to assess the accuracy of SSVEP detection. The accuracy was computed as a number of properly detected frequencies divided by number of loop iterations. The number of loop iterations represents the number of segments in which the data was processed. Appropriate flickering frequency was detected within the classification tolerance ± 1 Hz.

The results show that the CCA method was more accurate than the CWT, with median accuracy of 85 %. The CWT method reached median accuracy of 77 %. The CCA method was presented in 2006 for the first time [17]. Many studies investigated and applied this method for the SSVEP detection [7], [9], [17]. The CCA method usually provides better results in comparison with other methods.

The application of CWT for SSVEP detection was presented and compared with application of the power spectral density in 2010 [16]. The comparison of individual wavelets, the evaluation of accuracy and the evaluation of time resolution were performed for both methods. In our implementation the CWT method had higher computational time consumption compared to the CCA. The computational requirements could be decreased by code optimization and reduction of scale which would contain lower number of frequency bands. The SSVEP frequency is then detected in the nearest frequency band.

5 Conclusion

In this paper, we compared two methods for SSVEP detection. Results show that the detection accuracy is 77 %, using a CWT method. The CCA with accuracy 85 % is the most successful method. In order to further improve detection accuracy, we need to take into account harmonics of flickering frequency and implement it in a suitable classification technique. This is a future task.

Acknowledgement

The work and the contributions were supported by the project SP2013/35 “Biomedical engineering systems IX” and TACR TA01010632 “SCADA system for control and measurement of process in real time”.

References

1. J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Electroenceph. Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, June 2002.
2. Schalk, G.; McFarland, D.J.; Hinterberger, T.; Birbaumer, N.; Wolpaw, J.R., "BCI2000: a general-purpose brain-computer interface (BCI) system," *Biomedical Engineering, IEEE Transactions on*, vol.51, no.6, pp.1034-1043, June 2004.
3. Han-Jeong Hwang, Jeong-Hwan Lim, Young-Jin Jung, Han Choi, Sang Woo Lee, Chang-Hwan Im, Development of an SSVEP-based BCI spelling system adopting a QWERTY-style LED keyboard, *Journal of Neuroscience Methods*, Volume 208, Issue 1, 30 June 2012, Pages 59-65, ISSN 0165-0270.
4. F. Beverina, G. Palmas, S. Silvoni, F. Piccione, and S. Giove, "User adaptive BCIs: SSVEP and P300 based interfaces," *Psychology Journal*, vol. 1, pp. 331–354, 2003.
5. C. S. Herrmann, "Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena," *Experimental Brain Research*, vol. 137, no. 3-4, pp. 346–353, 2001.
6. M.A. Pastor, J. Artieda, J. Arbizu, M. Valencia, J.C. Masdeu, "Human Cerebral Activation during Steady-State Visual-Evoked Responses," *The Journal of Neuroscience*, Vol.23 Issue 37, pp.11621-11627, 2003.
7. Qingguo Wei; Meixia Xiao; Zongwu Lu; , "A Comparative Study of Canonical Correlation Analysis and Power Spectral Density Analysis for SSVEP Detection," *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2011 International Conference on*, vol.2, no., pp.7-10, 26-27 Aug. 2011.
8. Punsawad, Yunyong; Wongsawat, Yodchanan; , "Multi-command SSVEP-based BCI system via single flickering frequency half-field stimulation pattern," *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, vol., no., pp.1101-1104, Aug. 30 2011-Sept. 3 2011.
9. Guangyu Bin.; Zhonglin Lin.; Xiaorong Gao.; Bo Hong.; Shangkai Gao, "The SSVEP topographic scalp maps by Canonical correlation analysis," *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, vol., no., pp.3759-3762, 20-25 Aug. 2008.
10. Yu-Te Wang; Yijun Wang; Chung-Kuan Cheng; Tzyy-Ping Jung; , "Measuring Steady-State Visual Evoked Potentials from non-hair-bearing areas," *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, vol., no., pp.1806-1809, Aug. 28 2012-Sept. 1 2012.
11. Hui Li; , "Complex Morlet wavelet amplitude and phase map based bearing fault diagnosis," *Intelligent Control and Automation (WCICA), 2010 8th World Congress on*, vol., no., pp.6923-6926, 7-9 July 2010.
12. BORG, Magnus. *Learning multidimensional signal processing*. Linköping: Univ, 1998. ISBN 91-721-9202-X.
13. J. G. Proakis, *Digital Communications*, 4th Edition, McGraw-Hill, 2001.
14. "Signal Processing Toolbox - Parametric Methods", [online]. Available: <http://radio.feld.cvut.cz/matlab/toolbox/signal/spectr9a.html> [Accessed Jan. 21, 2013].
15. M. H. Hayes. *Statistical Digital Signal Processing and Modeling*. New York: Wiley, 1996.
16. Zimu Zhang; Xiuquan Li; Zhidong Deng; , "A CWT-based SSVEP classification method for brain-computer interface system," *Intelligent Control and Information Processing (ICICIP), 2010 International Conference on*, vol., no., pp.43-48, 13-15 Aug. 2010.
17. Zhonglin Lin; Changshui Zhang; Wei Wu; Xiaorong Gao; , "Frequency Recognition Based on Canonical Correlation Analysis for SSVEP-Based BCIs," *Biomedical Engineering, IEEE Transactions on*, vol.53, no.12, pp.2610-2614, Dec. 2006.

RFID Expert Database

Lucie Švecová

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
va1305@vsb.cz

Abstract. This paper introduces a RFID expert database. RFID and RTLS systems are currently a big help with the cost reduction not only in logistics, but also in other sectors of industry and services. Because of the wide range of use of these technologies, which moreover still evolving, there is an effort to facilitate of orientation in these modern methods of identification for interested. The database described the parameters of the components of RFID systems, which have been analyzed as important when choosing a solution. One of the parts of the database is also preview of finished and functional projects.

1 Introduction

At the Technical University of Ostrava in cooperation with Gaben s.r.o. is created the RFID expert database. This database will provide a summary about possibilities of using radio frequency identification (RFID). Along with an overview of available products, there is an introduction of ready-made applications. RFID is experiencing a boom and covers more areas of industry and public life. It is difficult to navigate in such a large amount of information for prospective customers from the very different fields of industry and services. The database will guide them on the way to choose the most suitable RFID for their intended application.

2 RFID areas of use

Radio frequency identification is a modern method of contactless identification. Initial RFID applications were used mainly in logistics as a complement or replacement of the barcode. RFID gradually expanded from purely logistical applications, such as tracking of goods and material flow, to other areas.

Today, it is often used for security applications. Access cards or chips for opening doors in businesses and hotels as well as residential buildings are commonly used. The Blood Center of the University Hospital of Ostrava has installed RFID locks since 1999. This prevents donors and patients as well as unauthorized personnel to access enclosed workplaces.

RFID is also used in case of identifying a large number of people in a short time, such as cultural and sports events. RFID tag is part of the ticket. This year's four-day

Bonnaroo Music and Art Festival in Tennessee attracted around 80,000 visitors, who instead of tickets received bracelet – high frequency RFID tag. In addition to the ticketless entry, RFID wristband was used to share "likes" and photos on social networks. [1]

Beside replacement of paper tickets, RFID is used as an identifier for participants in mass sports events such as metropolitan marathons. On the famous international marathon in New York City in 2011 was presented system that brings spectators closer to competitors. The solution uses data transmitted from stationary RFID readers placed along the race route and the information transmitted from the runner's GPS module smartphones. Runners have smartphones mounted on an arm that does not hinder them from running. Spectators can watch the continuous time of each competitor along with the current athlete's location on the track. Also athletes have access to the information during competition through their smartphones. [4]

RFID finds its place also in agriculture, where helps to save working time and reduce costs thanks to the rapid identification. Field workers on Bodega Norton Winery were given color- and shape-coded plastic and aluminum chips for each bin of grapes they picked.

"At the end of each week, their compensation was calculated based on the number and types of chips they turned in to vineyard managers. During any given week, thousands of chips would be distributed and collected from all workers. But the process was complicated, the company reports, because workers often lost the chips, and tracking the payments on paper proved to be error-prone and unreliable. What's more, employees and supervisors were spending one day each week collecting chips and processing payments, instead of harvesting grapes. To wring out those inefficiencies, Bodega Norton introduced a radio frequency identification solution in 2011 that provides workers with an extra picking day each week, and improves visibility into operations." [2]

"This solution has drastically improved Bodega Norton's productivity through more efficient use of resources," says Pablo Minatelli, the vineyard's manager and project leader. "We have also discovered that the new system generates more detailed, real-time information that can be used to guide better management decisions." [2]

A separate chapter for the use of RFID is health care. Especially in hospitals can find tasks for most of the above mentioned applications. A common application is stocktaking. Also contactless payment transactions are carrying on by Near Field Communication, which is based on RFID.

Besides the identification itself, there is an increasing demand for Real-time locating system (RTLS). RTLS is aimed to asset tracking or monitoring people. Among the asset tracking belong monitoring of equipment, tools, and carts in factories, warehouses and on industrial construction sites.

"On average, each site tracks between 10,000 and 100,000 critical items, such as fabricated pipe spools, valves and cable. At one such location, the operator has reduced its staff by 77 percent, by decreasing the amount of labor required to search for lost materials." [3]

"Regarding the oil-sands industry in Canada, the need to reduce the number of man-hours employees spend looking for missing materials is paramount. Highly paid and skilled workers can often spend hours searching for something that may or may not be located onsite. Not only does this consume man-hours, but a missing item can

also result in a construction delay. For that reason, materials are frequently reordered simply to reduce the need to search for them.” [4]

Monitoring people are focused to secure an access to the workplace and a labour safety. Active tags can recognize the unusual position of a person, e.g. lying person, and trigger the alarm. In the health services monitoring people have a future too, especially in the emergency intake, which in the case of a larger number of patients, long inactivity by the patient will be reported to the medical staff.

3 RFID Expert database

RFID systems consist of readers and tags. Readers transmitted at regular intervals an electromagnetic signal of a certain frequency. Tag located by reader receives signal sends back a response with identification data. Tag consists of a chip, antenna and packing. The chip is core of the tag. It communicates with reader and carries required data. The chip with the antenna imbedded on the substrate create inlay. Inlay in the packing form tag. The packing and thus form of the tag is selected according to application. They can be plastic cards, wristbands, labels, capsules for injection under the skin but also the packing that can withstand extreme temperatures and chemicals and ensure the operability of the inlay.

The initial implementation of RFID expert database, created on the open source database MySQL, is proceed in cooperation with Gaben s. r. o. The first task was analyzing, which kind of information about RFID components are indispensable for the search, and which are redundant. The most important information is:

- operating frequency
- communication protocol
- tag dimensions
- read range
- memory size
- application
- life time

As noted above, RFID systems consist of particular elements - tags and readers. Tags contain inlays and inlays contain chips. These four components are inserted into the database with basic technical parameters.

Chip, heart of the tag and carrier of information, is described mainly by communication protocol and operating frequency. Operating frequency indicates the frequency band in which the chip operates. Figuratively it also means throughput of electromagnetic signal in substances. Another indication is TID memory size, which is a unique identifier. There is also reported memory size and its distribution. Interesting in RFID chip will be the size of EPC memory and user memory, as well as the possibility of securing the tag with access password or possibility of destroying - kill command. Manufacturers provided chips with additional technologies, which is also a qualitative description of the chip and have a place in the database.

d_inlay_model_name	cat_name	cat_description	value	post	val_type
416	34 ALN-9634	Pracovní teplota	Maximální	70 °C	numeric
417	34 ALN-9634	Pracovní teplota	Minimální	-40 °C	numeric
418	34 ALN-9634	Provedení	Typ inlay	mokrý	numeric
419	34 ALN-9634	Rozměry	Anténa deska	44 mm	numeric
420	34 ALN-9634	Rozměry	Anténa síťka	46 mm	numeric
421	34 ALN-9634	Rozměry	Inlay síťka	47 mm	numeric
422	34 ALN-9634	Rozměry	Inlay síťka	51 mm	numeric
423	35 ALN-9640	Cykly	Zápis	100000	numeric
424	35 ALN-9640	Pracovní teplota	Maximální	70 °C	numeric
425	35 ALN-9640	Pracovní teplota	Minimální	-40 °C	numeric
426	35 ALN-9640	Provedení	Typ inlay	mokrý	numeric
427	35 ALN-9640	Rozměry	Anténa deska	94.8 mm	numeric
428	35 ALN-9640	Rozměry	Anténa síťka	8.1 mm	numeric
429	35 ALN-9640	Rozměry	Inlay síťka	98.2 mm	numeric
430	35 ALN-9640	Rozměry	Inlay síťka	12.3 mm	numeric
431	36 ALN-9654	Cykly	Zápis	10000	numeric
432	36 ALN-9654	Pracovní teplota	Maximální	70 °C	numeric
433	36 ALN-9654	Pracovní teplota	Minimální	-40 °C	numeric
434	36 ALN-9654	Provedení	Typ inlay	mokrý	numeric
435	36 ALN-9654	Rozměry	Anténa deska	93 mm	numeric
436	36 ALN-9654	Rozměry	Anténa síťka	10 mm	numeric

Fig. 1 Initial implementation of database – inlay

Inlay is a "semiproduct" of the tag. It contains a chip and antenna, substrate pad is made of plastic, most often PET, PVC or PC. Inlay is characterized by dimensions of the antenna, which indicate the final size of inlay in the so-called "dry inlay", and by the die-cut dimensions of "wet" version. Inlays are different mainly in the size and the shape of antenna, depending on the type of application and the frequency of the chip. Next property is operating temperature, which may differ from the operating temperature of tag, in which inlay is embedded.

Tag is described by dimensions, operating temperature, life time, material and kind of application, which primary is suited for. Next, there are passive and active tags. Passive tags are common, active tags may have longer read range or can be equipped with environment sensors.

Readers have also communication protocol and operating frequency such a main parameters. Readers are differed as stationary, handheld and mobile devices. Describing properties are number, size and type of antenna, reading/writing time and also possibility of multiple reading.

Conclusion

RFID and RTLS are modern technologies bringing new qualitative characteristics that can greatly enrich contemporary society. Perhaps the greatest benefit of RFID and RTLS technology is recorded in logistics. The increasing demands of legislation can be manage just with these technologies, particularly in the backward traceability. The use of RFID and RTSL tags in a transport services can significantly reduce unnecessary logistical movements and optimize traffic. Expert database is created just because of the RFID and RTLS expansion. It is for those, who are interested in these technologies for improving their businesses.

References

1. Swedberg, C.: RFID Performs a Bigger Role at Bonnaroo Festival. RFID Journal
2. Gambon, Jill.: Argentine Winery Harvest Crop With RFID. RFID Journal
3. Swedberg, C.: RFID Tracks Assets at Canada's Oil Sands. RFID Journal
4. <http://www.rfid-epc.cz/clanky/reseni/rfid-na-bezecke-trati-a2507167>

Using Mathematical Background for Determining the Temperature Distributions and Corresponding Parameters inside a Rock Massif

Petr Vojčínák

Department of Cybernetics and Biomedical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
petr.vojcinak@vsb.cz

Abstract. Geothermal resources are economically used for parts of geothermal energy, whereas its resources are utilized nowadays and specified via a number of boreholes, and geophysical and geochemical data. In case of potential geological localities, the resources could be used, but relevant data have not been acquired to evaluate them quantitatively. So this paper deals with some usage of a full-scale *in-situ* TRT experiment's framework for detailed determining temperature conditions inside borehole heat exchangers, namely at considering some blocks including information on geological and geographical conditions, automatized measurements of undisturbed/disturbed temperature profiles, and acquired TRT data evaluations. This paper deals with illating, detailed describing, and summarizing some temperature distributions, based on Kelvin line-source theory, which is closely related to the first level of evaluating the TRT data.

Keywords: Borehole, Borehole heat exchanger (BHE), Cauchy problem, Fourier partial differential equation, Heat transfer, ILS/FLS model, Kelvin line-source theory, Mathematical modelling, Rock massif, Thermal conductivity, Thermal resistance, TRT experiment.

1 Introduction

1.1 Features of Thermal Response Test (TRT)

Thermal response test (TRT abbreviation) or geothermal response test (GRT or GeRT abbreviation) is a field method, how to investigate some thermo-physical properties of a BHE, which is drilled into a soft/hard rock massif, and its surroundings. The BHE typically contains one or more loops (PE pipes; e. g. types of a single U-tube with one pair of pipes or one collector circuit, or a double U-tube with two pairs of pipes or two collector circuits), where a heat carrier fluid is pumped into the collector circuit at a constant rate (typically a constant mass-flow, M_j in $\text{kg}\cdot\text{s}^{-1}$). This fluid – in case of our TRT experiments, pure water is usually used – may be heated by a TRT device's heater at a constant injected heating power (Q ; in W), whereas quantities of an inlet fluid temperature (T_{jIN} ; in K or $^{\circ}\text{C}$), an outlet fluid temperature (T_{jOUT} ; in K or $^{\circ}\text{C}$), a

volume-flow (Q_V ; in $\text{m}^3\cdot\text{s}^{-1}$), a pressure (p ; in Pa), and other electric and/or non-electric ones are measured and logged continuously [1, 2].

The U-tube look/loops, through which the heat carrier fluid circulated, is/are inserted inside the borehole to the same depth as the BHE, planned for a site. To provide a constant heat flux (q , in $\text{W}\cdot\text{m}^{-2}$) to the ground, a fluid flow rate (also the mass-flow) inside the loop/loops, and a temperature difference between the inlet and outlet fluid temperatures must be constant during a TRT experiment. [2]

For determining some essential ground thermal properties – typically a thermal conductivity (λ ; in $\text{W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$ or $\text{W}\cdot\text{m}^{-1}\cdot^\circ\text{C}^{-1}$) and a thermal borehole resistance (R_B ; in $\text{K}\cdot\text{m}\cdot\text{W}^{-1}$ or $^\circ\text{C}\cdot\text{m}\cdot\text{W}^{-1}$) – from output TRT data [2], different kinds of analytical and numerical methods have been developed yet. The BHE, which consists of two pipes or one collector circuit, separated by a filling material (also a borehole grout), can be modelled as a heat source in the form of a line (line-source theory; ILS/FLS model by Carslaw and Jaeger, 1959) or a cylinder (cylinder-source theory; ICS/FCS model by Ingresoll et al., 1954) [3].

A standard TRT experiment is based on acquiring the data, used for design and model processing of some larger BHE field (e. g. a research polygon with one or more BHEs). Using the TRT experiment, it is also possible to evaluate a technical integrity of these BHEs. In case of a conventional TRT experiment's evaluation approach, based on *the line-source theory*, it is required long-term test durations (t_{test} ; in s), typically taking 48 hours or 72 hours, to obtain some quasi-steady state (which is time-independent) conditions. In case of applying *the cylinder-source theory*, a precision of the evaluation is enhanced and the test durations are reduced, because unsteady state (which is time-dependent) is evaluated. [2, 3]

2 Mathematical Background of the TRT

In this section, a simplified mathematical background, used for evaluating results in terms of these TRT tests, is typically based on *the Kelvin line-source theory*. At first, let us consider *the Fourier equation of heat transfer* (i. e., it is the second order partial linear differential equation of parabolic type) in this generalized form, thus [4]:

$$\frac{\partial w}{\partial t} - L_{\hat{x},t}[w] = \Phi(\hat{x}, t) \quad (1)$$

$$t = 0: \quad w = f(\hat{x})$$

Based on the form of Eq. (1), we assume these conditions and simplifications, thus [4]:

- two spatial coordinates in the Cartesian coordinate system (i. e., at $n = 2$),
- one continuous time coordinate, leading to a transient solution of this problem,
- no internal heat source,
- isotropic thermal conductivity,

- initial and boundary conditions of the Cauchy problem,
- a fundamental solution of the Cauchy problem.

Equations (2) ~ (4) summarize these conditions as follows, thus [4]:

$$\begin{aligned} \hat{x} &= \{x, y\} \in \mathfrak{R}^2 && \text{(m)} \\ w &\equiv w(\hat{x}, t) \equiv T(x, y, t) && \text{(K)} \\ L_{\hat{x},t}[w] &\equiv \frac{\lambda}{\rho \cdot c} \cdot \left\{ \frac{\partial^2 T(x, y, t)}{\partial x^2} + \frac{\partial^2 T(x, y, t)}{\partial y^2} \right\} && \text{(K} \cdot \text{s}^{-1}) \\ a &\equiv \frac{\lambda}{\rho \cdot c} = \frac{\lambda_x}{\rho \cdot c} = \frac{\lambda_y}{\rho \cdot c} = \frac{\lambda_z}{\rho \cdot c} && \text{(m}^2 \cdot \text{s}^{-1}) \\ &\Phi(\hat{x}, t) \equiv 0 && \text{(K} \cdot \text{s}^{-1}) \end{aligned} \tag{2}$$

at (initial and boundary conditions; in kelvins and meters, respectively)

$$\begin{aligned} \text{IC: } t = 0: \quad w &= f(x, y) && \text{(K)} \\ \text{BC: } \quad -\infty < x < +\infty && \text{(m)} \\ \quad -\infty < y < +\infty && \end{aligned} \tag{3}$$

and (the fundamental solution of the Cauchy problem; non-dimensional)

$$g(x, y, \xi, \eta, t, \tau) = \frac{1}{4 \cdot \pi \cdot a \cdot (t - \tau)} \cdot \exp \left[-\frac{(x - \xi)^2 + (y - \eta)^2}{4 \cdot a \cdot (t - \tau)} \right] \tag{4}$$

and also (the transient solution of this problem; at zero initial temperature, at $\tau = 0$ and in kelvins)

$$T(x, y, t) = \frac{1}{4 \cdot \pi \cdot a \cdot t} \cdot \iint_{\mathfrak{R}^2} f(\xi, \eta) \cdot \exp \left[-\frac{(x - \xi)^2 + (y - \eta)^2}{4 \cdot a \cdot t} \right] \cdot d\xi \cdot d\eta \tag{5}$$

Equation (5) is related to a distribution of so-called *instantaneous point source* of strength $\Phi \cdot d\zeta$ at ζ along the z -axis (in meters). In this case, the quantity of heat liberated per unit length of the line is equalled to $\Phi \cdot c \cdot \rho$ (in joules per meter). [5]

In case of so-called *continuous line source* (i. e., continuous denotes the transient solution depending on the continuous time t or t -coordinate), we suppose heat to be liberated at rate $\Phi(t) \cdot c \cdot \rho$ (in joules per meter per second or in watts per meter) per unit time per unit length of a line, parallel to the z -axis and through the point $[\xi, \eta]$. If supply of heat starts at $t = 0$ (when the solid is at non-zero temperature; in kelvins),

then the temperature at the t -coordinate is given by this useful formula, thus (in kelvins): [5]

$$T(x, y, t) = T(x, y, 0) + \frac{1}{4 \cdot \pi \cdot a} \cdot \int_0^t h(\tau) \cdot \frac{1}{t - \tau} \cdot \exp\left[-\frac{(x - \xi)^2 + (y - \eta)^2}{4 \cdot a \cdot (t - \tau)}\right] \cdot d\tau \quad (6)$$

If $\Phi(t) = q = konst.$ (in kelvins per square meter per second) and $T(x, y, 0) = f(x, y) = T_0$ (in kelvins), then Eq. (6) can be written in these possible forms including so-called *exponential integral function* $Ei(\cdot)$, thus (in kelvins): [5]

$$T(x, y, t) = T_0 + \frac{q}{4 \cdot \pi \cdot a} \cdot E_1\left[\frac{(x - \xi)^2 + (y - \eta)^2}{4 \cdot a \cdot t}\right] \quad (K)$$

$$T(x, y, t) = T_0 - \frac{q}{4 \cdot \pi \cdot a} \cdot Ei\left[-\frac{(x - \xi)^2 + (y - \eta)^2}{4 \cdot a \cdot t}\right] \quad (K) \quad (7)$$

$$T(x, y, t) = T_0 - \frac{q}{4 \cdot \pi \cdot a} \cdot \left[\gamma + \ln\left|-\frac{R^2}{4 \cdot a \cdot t}\right| + \underbrace{\sum_{k=1}^{\infty} \frac{\left(-\frac{R^2}{4 \cdot a \cdot t}\right)^k}{k \cdot k!}}_{\text{high-order terms}} \right] \quad (K)$$

where (by using the L'Hospital rule and Euler-Mascheroni constant, and the Fourier number, respectively)

$$-\gamma = \lim_{\beta \rightarrow 0} \frac{\beta \cdot \Gamma(\beta) - 1}{\beta} = \lim_{\beta \rightarrow 0} \frac{1 \cdot \Gamma'(1) - 0}{1} = \Gamma'(1) \approx 0,577215665 \quad (1)$$

$$R = \sqrt{(x - \xi)^2 + (y - \eta)^2} \neq 0 \quad (m) \quad (8)$$

$$Fo_R = \frac{a \cdot t}{R^2} = \frac{\lambda}{c \cdot \rho} \cdot \frac{t}{R^2} \quad (1)$$

$$E_1(\cdot) = E_{1-\beta}(\cdot) \Big|_{\beta=0} \quad (1)$$

If we consider items of Eq. (8), a transformation of Eq. (1) and Eq. (2) into the cylindrical coordinate system $[r, \varphi]$, and an axi-symmetrical problem, then we reduce this two-dimensional problem to one-dimensional problem with the r -coordinate only, and get these forms of a particular solution (at $\beta = 0$, non-dimensional integration constant), thus (in kelvins):

$$T(r,t) = A + B \cdot \left[\Gamma(\beta) - \gamma \left(\beta, \frac{r^2}{4 \cdot a \cdot t} \right) \right] \quad (\text{K})$$

$$T(r,t) = A + B \cdot E_1 \left(\frac{r^2}{4 \cdot a \cdot t} \right) \quad (\text{K})$$

(9)

$$T(r,t) = T_0 + \frac{q}{4 \cdot \pi \cdot a} \cdot \left[-\gamma + \ln \left(\frac{4 \cdot a \cdot t}{r^2} \right) + \underbrace{\sum_{k=1}^{\infty} (-1)^{k+1} \cdot \frac{\left(\frac{r^2}{4 \cdot a \cdot t} \right)^k}{k \cdot k!}}_{\text{high-order terms}} \right] \quad (\text{K})$$

If we compare the result forms of Eq. (7) and Eq. (9), then they are formally same and are valid for small time evaluation (i. e., transient behaviour). If the high-order terms are ignored, then Eq. (7) and Eq. (9) are valid for large time evaluation (i. e., quasi-stationary/stationary behaviour), so these TRT tests require long-time duration (t_{test} ; in s). At practical evaluating the TRT data, we also consider so-called *Lower Time Criterion* (at $t = t_b$ and $r = r_B$; in s and m, respectively), closely related to calculating relative errors (δ ; in %).

If Eq. (9) practically satisfies values of a borehole radius (r_B ; in m), a borehole depth (h_B ; in m), and average values of temperature of a heat-carrier fluid ($\overline{T_f(t)}$; in K), internal thermal resistivity of the borehole ($\overline{R_B(t)}$; in $\text{K} \cdot \text{m} \cdot \text{W}^{-1}$), temperature of a borehole wall ($\overline{T_B(t)}$; in K), temperature of a surrounding rock massif ($\overline{T_G(t)}$; in K), and no thermal resistance at a borehole wall/rock massif boundary (i. e., $\overline{T_B(t)} = \overline{T_G(t)}$ and $r = r_B$), then this equation can be written in these forms (at quasi-stationary behaviour and $q_{h_b} = Q/h_B$; in $\text{W} \cdot \text{m}^{-1}$), thus (in kelvins):

$$\left[T_f(t) - \overline{T_f(t=0)} \right] - \overline{T_B(t)} = \overline{R_B(t)} \cdot q_{h_b} \quad (\text{K})$$

$$\left[T_f(t) - \frac{T_{f_{IN}}(0) + T_{f_{OUT}}(0)}{2} \right] - \frac{Q}{4 \cdot \pi \cdot \lambda \cdot h_B} \cdot \left[-\gamma + \ln \left(\frac{4 \cdot a \cdot t}{r_B^2} \right) \right] = R_B \cdot \frac{Q}{h_B} \quad (\text{K}) \quad (10)$$

Equation (10) is valid for average value between warmer inlet (i. e., to an U-tube) and cooler outlet (i. e., from the U-tube). This value is related to a fluid temperature (i. e., $T_f(t)$, $T_{f_{IN}}(t)$ and $T_{f_{OUT}}(t)$; in K), thermal diffusivity (i. e., a , a_{IN} and a_{OUT} ; in $\text{m}^2 \cdot \text{s}^{-1}$), thermal conductivity (i. e., λ , λ_{IN} and λ_{OUT} ; in $\text{W} \cdot \text{m}^{-1} \cdot \text{K}^{-1}$), and the internal thermal resistivity of the borehole (i. e., R_B , $R_{B_{IN}}$ and $R_{B_{OUT}}$; in $\text{K} \cdot \text{m} \cdot \text{W}^{-1}$), so we must calculate the practical results (i. e., inlet temperature curve $T_{f_{IN}}(t) = f_1(t)$ and/or outlet temperature curve $T_{f_{OUT}}(t) = f_2(t)$) for inlet and outlet separately.

3 Conclusion

The first part of this paper is centred on a general description of the TRT approach, based on the ILS/FLS model and the *in-situ* method in this case, and the summarization of all the localities (their specific names have not been mentioned due to commercial obligations), where the TRT experiments have been realized by our TRT device of UBeG.

The second part of this paper is centred on mathematical formulating this two-dimensional problem (i. e., the second order souceless PDE of parabolic type, IC and BC inside this domain), focused on the Cauchy problem, when the boundary conditions are valid for infinite intervals of both spatial coordinates. In the Cartesian coordinate system, the elementary forms of desired temperature distributions (as particular solutions) were introduced in Eq. (6) and then in Eq. (7); in the cylindrical coordinates, the temperature distributions were introduced in Eq. (9). Finally, Eq. (10) transforms theoretical mathematical formulae into the practical ones for the first level of evaluating the TRT data.

Aknowledgement

This work is supported by the project SP2013/168, named “Methods of Acquisition and Transmission of Data in Distributed Systems”, of Student Grant System (VŠB – Technical University of Ostrava).

This research is supervised by doc. Ing. Jiří Koziorek, Ph.D.

References

1. P. Vojčínák, R. Hájovský, J. Koziorek, “Using ILS Mathematical Model and Measurements for Detailed Evaluating Results of Thermal Response Tests”, In ICCAS 2012: Proceedings of the 12th International Conference, Jeju Island (KR), 2012.
2. P. Vojčínák, R. Hájovský, J. Koziorek, “Using TRT for Determining the Essential Parameters of Rock Massif”, In PDES 2012: Proceedings of the 11th IEEE/IFAC International Conference, Brno (CZ), 2012.
3. P. Vojčínák, P. Pospíšil, R. Hájovský, J. Koziorek, “Using Improved Methodology for Investigating Results of Thermal Response Test Experiments”, In HTE 2012: Proceedings of the 10th WSEAS International Conference, Istanbul (TR), 2012.
4. A. D. Polyanin, “Handbook of Linear Partial Equations for Engineers and Scientists”, Boca Baton, Chapman and Hall/CRC (USA), 2002, ISBN 1-58488-299-9.
5. H. S. Carslaw, J. C. Jaeger, “Conduction of Heat in Solids”, Oxford, Clarendon Press (UK), 1959.

Risk Modeling of Accidental Events and Responses in Process Industries

Ondřej Grunt

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
ondrej.grunt@vsb.cz

Abstract. Modeling the risk to safety of personnel in offshore industry is often realized by the application of Event Trees. The risk is then defined as a product of event frequency and its consequences. This article presents a new possible approach to modeling the risk to safety of personnel by substituting the Event Tree model with Stochastic Petri nets model. Stochastic Petri nets model for small leak occurrence on an offshore platform is shown, based on realistic example from the offshore industry. The probabilities of fatalities were obtained from the simulation by using the Moca-RP software.

Keywords: Risk modeling; Stochastic Petri nets; Event Trees;

1 Introduction

Personnel in process industries is exposed to the risk of fatality. The risk of fatality is normally estimated by an application of Event Tree method. An event tree displays a sequence of events with their corresponding probabilities, resulting probabilities of fatalities are represented by the probability values assigned to the Event Tree leaves, each leaf representing a particular event scenario.

However, tracing contributions of fatality probabilities in an Event Tree is difficult. While it is possible to construct an Event Tree large enough to enable us to find the most important contributions, such an Event Tree would be very difficult to manage. Another limitation is that the Event Tree method is a steady-state method, therefore not suitable for modeling time-dependent processes.

In the offshore industry, fatality probabilities often depend on personnel reactions and their consequences. Due to the lack of aforementioned limitations bound with the Event Tree method, Stochastic Petri nets (SPN) were considered to be used to construct a suitable model of the risk to safety of personnel. As an initial step of the substitution of the Event Tree method in the risk modeling, steady-state representation of a time-dependent event was modeled using SPN. In this article, a small leak occurrence scenario was chosen as an example.

Probabilities of fatalities due to small leak occurrence obtained by the Moca-RP simulation of Stochastic Petri nets were compared to results obtained by the application of Monte Carlo method [1].

2 Petri nets

Petri nets is a modeling language for the description of parallel and distributed systems, designed by Carl Adam Petri. Petri net is a directed bipartite graph, its nodes are divided into two groups : *places* and *transitions*. Places are nodes representing current condition of the system (e.g. "power on", "power off") and are signified by circles, transitions are nodes representing events occurring in the system (e.g. "powering on", "powering off") and are signified by bars. Places and transitions are connected by directed arcs, however a place can only be connected via arc to a transition and not to another place (i.e. each event leads to a new condition of the system). Similarly, transition cannot be connected via arc to another transition.

Places in a Petri net may contain a number of *tokens*. Distribution of tokens over the places in Petri net is called a *marking*, which represents a current configuration of the net. Token proceeds to another place by firing an enabled transition. Transition is enabled only if the input place contains a sufficient number of tokens, these tokens are then transferred from the input place to an output place. Basic Petri nets model is nondeterministic, therefore any enabled transition may fire immediately after a sufficient number of tokens is present in the input place.

A Petri net can be defined as follows:

Definition 1. A *Petri net* is a four-tuple $PN = (P, T, F, M_0)$ where:

1. P is a set of **places**.
2. T is a set of **transitions**.
3. $F, F \subset (P \times T) \cup (T \times P)$ is a set of **arcs**.
4. M_0 is the **initial marking**.

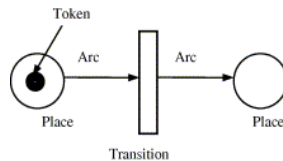


Fig. 1. Example of a simple Petri net.

However, for modeling the risk to safety of personnel, the nondeterminism of basic Petri nets is a significant issue. Every reaction of personnel to an occurring event during a scenario contributes to the probability of fatality. For this reason, a decision concerning which of the multiple transitions may fire must be confined to a set of carefully chosen laws.

2.1 Stochastic Petri nets

Stochastic Petri nets (further referred to as SPN) is a form of Petri nets in which a probabilistic delay is assigned to each transition. After the delay is over, transition is allowed to fire.

A Stochastic Petri net can be defined as follows:

Definition 2. *A Stochastic Petri net is a five-tuple $SPN = (P, T, F, M0, \Lambda)$ where:*

1. *P is a set of places.*
2. *T is a set of transitions.*
3. *$F, F \subset (P \times T) \cup (T \times P)$ is a set of arcs.*
4. *$M0$ is the initial marking.*
5. *Λ is the array of transition firing rates λ .*

3 Application of SPN

The application in this article is based on a typical offshore hydrocarbon installation (well-described example of an offshore production installation can be seen in [1]).

Risks from potential small leaks of the produced hydrocarbons were used as an example for the application. Small leaks were given precedence over the larger leaks due to substantially higher estimations of the frequencies of their release.

3.1 Small Leak scenario

After a small leak occurrence, personnel is alarmed and starts escaping. On the installation, there are primary, secondary and tertiary evacuation routes, each leading to lifeboats or life-rafts. In case of damage or blockade of all evacuation routes, personnel evacuates to the sea. In any case, escaped personnel is then gathered by a standby vessel. Fatality in this scenario is a direct result of the loss of probability of evacuation.

However, there is always a possibility that the leak may ignite. Consequences of the ignition depend on time it occurs. Immediate ignition results in a jet fire, which may cause fatality to the surrounding personnel, while delayed ignition may result in an explosion, possibly damaging or blocking the evacuation routes. In any event, resulting fire may escalate outside the zone, damaging the evacuation routes in the process. Any of these possible events strongly contribute to the probability of fatality.

3.2 SPN model of Small Leak scenario

As aforementioned in the beginning of this article, MOCA-RP simulation software was used for creating the model of the Small Leak scenario. In addition to the standard SPN elements, variables declared by the user may be utilized,

setting values is done by firing a transition with corresponding calling function, or may be used in place of a random value for determining delay of a transition. Another useful feature is a possibility to create *repeating places*, a set of pointers to a given place, improving lucidity of the model.

All used transitions (with the exception of the last pair of transitions) utilize exponential firing law with either fixed or variable-based rate. Reader should be reminded that in this article, the SPN model is used as a substitution of the Event Tree model.

Presented model of the Small Leak scenario is divided into several blocks:

- Main block
- Escalation blocks (3 in total)
- Evacuation blocks (4 in total)

The purpose of the Main block (Figure 2) is to determine whether a small leak occurrence ends with a fatality or not. Each token¹ in place *SmallLeak* represents one small leak occurrence on the installation. As aforementioned, the most important possible event is an ignition of the leak, this is resolved by the first pair of transitions. If the ignition occurred, its time is then determined by another set of transitions.

Purpose of the Evacuation blocks (Figure 3 on the right) is to alter the probability of fatality, alteration is based on availability of the lifeboats. Evacuation blocks are chosen in a similar manner as Escalation blocks with the addition of a separate block for a scenario, in which the leak does not ignite.

When a token arrives in place *ZoneProgressed*, transitions determining the success of the evacuation are enabled. Firing any of these transitions results in a transfer of token to a place according to the outcome of the evacuation (e.g. if the scenario ended with a fatality, token arrives in place *Fatality*).

The last pair of transitions provides cumulative frequencies of the fatalities and resets values of used variables. Another token then proceeds through the net, until all tokens are contained by place *End*.

4 Results

For computation of the probabilities of fatalities, MOCA-RP computing module was used. Four cases were distinguished by a nature of the ignition. For comparison with the results presented in [1] by application of the Event Tree method, same sample was chosen consisting of 609 occurrences of hydrocarbon small leaks.

As could be expected, delayed leak ignition is the most dangerous event, resulting in an explosion and thus strongly contributing to the loss of possibility of evacuation. This expectation is in accordance with presented results in Table 1.

¹ In MOCA software, tokens are called "Jets". Number assigned to a place is identifier for the software algorithms.

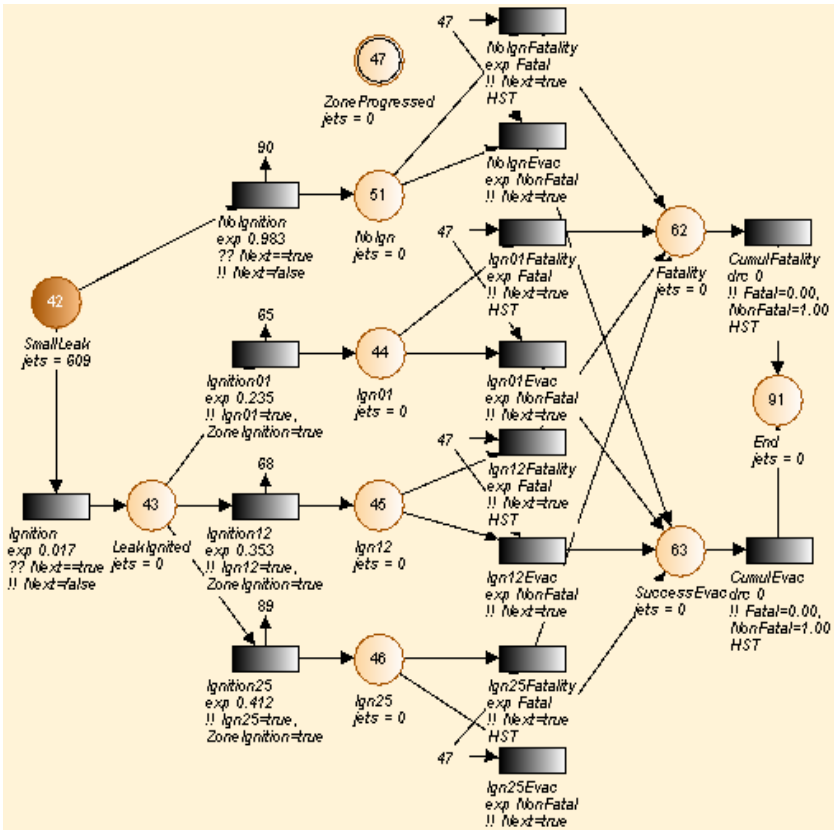


Fig. 2. SPN model : Main block.

Case	SPN model	ET model
1 (No ignition) :	$1.24e - 5$	$9.8e - 6$
2 (Ignition between 0 and 1 minute) :	$2.65e - 4$	$2.6e - 4$
3 (Ignition between 1 and 2 minutes) :	$8.51e - 5$	$8.5e - 5$
4 (Ignition between 2 and 5 minutes) :	$1.08e - 3$	$1.1e - 3$
Total :	$1.444e - 3$	$1.433e - 3$

Table 1. Resulting probabilities of fatalities obtained by the SPN and the ET model.

As can be seen in Table 1, the difference between the probabilities of fatalities in every case are minimal, with the most notable (albeit small) difference between the no ignition cases. Thus the results of the SPN method and Event Tree method are very close, with the cumulative probabilities of fatality differing by a margin of $1.1e - 5$.

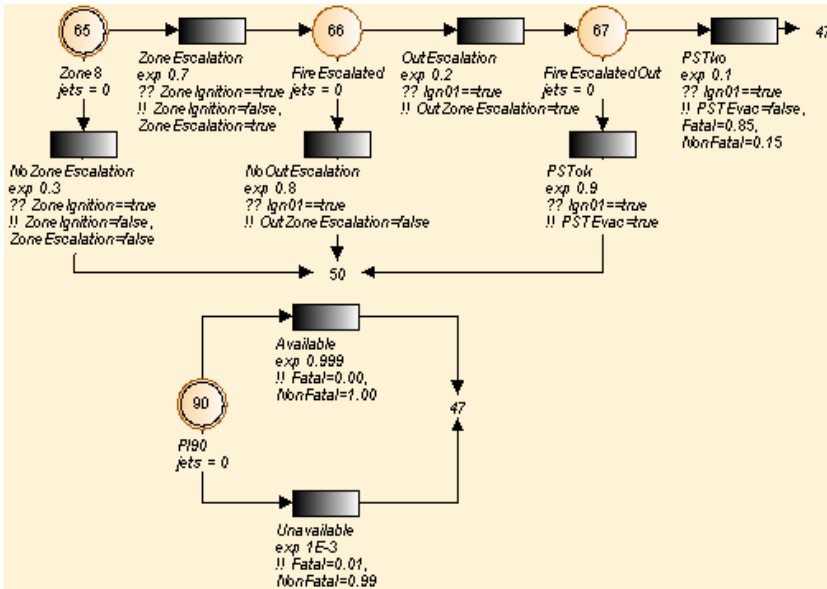


Fig. 3. SPN model : Escalation block and Evacuation block.

5 Conclusion

In this article, a new approach for modeling risk to safety of personnel in process industries was presented in the application of SPN. Benefits of SPN over normally used Event Trees were discussed and SPN formalism was presented.

Model representing Small Leak scenario on a typical offshore installation was used to compute the probabilities of fatalities and results were compared with those obtained by Event Tree model. As the differences between computed probabilities were very low, SPN can be considered a suitable substitution for Event Trees in process industry. While this paper focuses on the application of SPN as a steady state method of modeling, it can be used for modeling time-dependent processes. Proposed further work in this area is construction of time-dependent SPN model of the Small Leak scenario.

References

1. Radim Briš, Sava Medonos, Chris Wilkins, Adam Zdráhala, *Time-Dependent Risk Modeling of Accidental Events and Responses in Process Industries*, Proceedings of Esrel 2012, Helsinki, ISBN 978 162 276 4365.
2. Jean-Pierre Signoret, Yves Dutuit, Pierre-Joseph Cacheux, Stéphane Collas, Cyrille Foleaux, Philippe Thomas, *Reliability block diagrams driven Petri Nets*, Reliability Engineering and System Safety 113 (2013) 61–75.
3. M. Ajmone Marsan, Gianfranco Balbo, Gianni Conte, Susanna Donatelli, Giuliana Franceschinis, *Modeling with Generalised Stochastic Petri Nets*, Universita degli Studi di Torino, 1994.

Efficient Implementation of the TFETI Coarse Problem Solution in the FLLOP library

Václav Hapla

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
vaclav.hapla@vsb.cz

Abstract. FETI domain decomposition methods blend iterative and direct solvers. The original problem is reformulated using the dual theory of convex programming and the resulting dual problem is solved iteratively using the conjugate gradients algorithm. In each iteration, auxiliary problems related to the application of an unassembled dual operator (stiffness matrix pseudoinverse action and coarse problem solution) are solved. This paper deals with a performance analysis of the FETI coarse problem solution using two strategies and two parallel direct solvers (MUMPS, SuperLU). For the numerical experiments, we used our FLLOP library.

Keywords: TFETI; coarse problem; direct solver; FLLOP.

1 Introduction

The class of methods called FETI (Finite Element Tearing and Interconnecting) turned out to be one of the most successful for parallel solution of elliptic partial differential equations arising from many engineering problems. They blend iterative and direct solvers. The dual problem is solved by the iterative solver, e.g. conjugate gradients (CG); in each iteration, the auxiliary problems related to the application of an unassembled system matrix (subdomain problems' solutions and projector application in dual operator) are solved.

The first auxiliary problem is the stiffness matrix's pseudoinverse application. It is parallelizable without any data transfers because of a nice block-diagonal structure. The second one is the coarse problem (CP) solution appearing in the application of the projector onto the kernel of so called natural coarse space matrix. The CP couples the subdomains and accelerates convergence. However, this problem does not possess such a nice structure suitable for parallel processing; some communication is needed in this case.

We strive to maximize the number of subdomains to reduce the sizes of the subdomain stiffness matrices and so accelerate their factorization and pseudoinverse applications. However, the CP gets bigger at the same time and can easily become a bottleneck. Here we will try to address this issue.

The codebase used for the experiments is the FLLOP library (FETI Light Layer on top of PETSc) [17]. It is our novel, not yet published software package for solution of Quadratic Programming (QP) problems that acts like an

extension of PETSc framework. PETSc (Portable, Extensible Toolkit for Scientific Computation) [15] is a suite of data structures and routines for the parallel solution of scientific applications modelled by PDE.

2 TFETI for elasto-static problems

The FETI-1 method is based on the decomposition of the spatial domain into non-overlapping subdomains that are "glued" by Lagrange multipliers, enforcing arising equality constraints by special projectors. The original FETI-1 method assumes that the boundary subdomains inherit the Dirichlet conditions from the original problem, so that the dimensions of subdomain stiffness matrices' kernels may vary. The basic idea of Total-FETI (TFETI) [10] is to keep all the subdomains floating and enforce the Dirichlet boundary conditions by means of a constraint matrix and Lagrange multipliers, similarly to the gluing conditions along subdomain interfaces. This simplifies implementation of the stiffness matrix pseudoinverse. The key point is the local stiffness matrices can be effectively regularized and their kernels are known a priori, have the same dimension and can be formed without any computation only based on mesh data[11–13].

Let us consider a partitioning of a global domain Ω into N_s subdomains Ω^s and denote by \mathbf{K}^s the subdomain stiffness matrix and by \mathbf{R}^s a matrix whose columns span the kernel of \mathbf{K}^s . Let \mathbf{B}^s be a matrix with values $-1, 0, 1$ describing the gluing of the subdomains and

$$\mathbf{K} = \text{diag}(\mathbf{K}_1, \dots, \mathbf{K}_s), \quad \mathbf{R} = \text{diag}(\mathbf{R}^1, \dots, \mathbf{R}^{N_s}), \quad \mathbf{B} = [\mathbf{B}^1, \dots, \mathbf{B}^{N_s}]. \quad (1)$$

Let N_p denote the primal dimension, N_d the dual dimension, N_n the null space dimension and N_c the number of cores available for our computation.

Let us apply the duality theory to the primal problem

$$\min \frac{1}{2} \mathbf{u}^T \mathbf{K} \mathbf{u} - \mathbf{u}^T \mathbf{f} \quad \text{s.t.} \quad \mathbf{B} \mathbf{u} = \mathbf{o} \quad (2)$$

and let us establish the following notation

$$\mathbf{F} = \mathbf{B} \mathbf{K}^\dagger \mathbf{B}^T, \quad \mathbf{G} = \mathbf{R}^T \mathbf{B}^T, \quad \mathbf{d} = \mathbf{B} \mathbf{K}^\dagger \mathbf{f}, \quad \mathbf{e} = \mathbf{R}^T \mathbf{f},$$

where \mathbf{K}^\dagger denotes a generalized inverse matrix satisfying $\mathbf{K} \mathbf{K}^\dagger \mathbf{K} = \mathbf{K}$ and \mathbf{G} is a so-called *natural coarse space matrix*. We obtain a new minimization problem

$$\min \frac{1}{2} \boldsymbol{\lambda}^T \mathbf{F} \boldsymbol{\lambda} - \boldsymbol{\lambda}^T \mathbf{d} \quad \text{s.t.} \quad \mathbf{G} \boldsymbol{\lambda} = \mathbf{e}. \quad (3)$$

Further the equality constraints $\mathbf{G} \boldsymbol{\lambda} = \mathbf{e}$ can be homogenized to $\mathbf{G} \boldsymbol{\lambda} = \mathbf{o}$ by splitting $\boldsymbol{\lambda}$ into $\boldsymbol{\mu} + \tilde{\boldsymbol{\lambda}}$ where $\tilde{\boldsymbol{\lambda}}$ satisfies $\mathbf{G} \tilde{\boldsymbol{\lambda}} = \mathbf{e}$ (e.g. $\tilde{\boldsymbol{\lambda}} = \mathbf{G}^T (\mathbf{G} \mathbf{G}^T)^{-1} \mathbf{e}$) which implies $\boldsymbol{\mu} \in \text{Ker } \mathbf{G}$. We then substitute $\boldsymbol{\lambda} = \boldsymbol{\mu} + \tilde{\boldsymbol{\lambda}}$, omit terms without $\boldsymbol{\mu}$, minimize over $\boldsymbol{\mu}$ and add $\tilde{\boldsymbol{\lambda}}$ to $\boldsymbol{\mu}$.

Finally, the equality constraints $\mathbf{G} \boldsymbol{\lambda} = \mathbf{o}$ can be enforced by the projector $\mathbf{P} = \mathbf{I} - \mathbf{Q}$ onto the null space of \mathbf{G} , where $\mathbf{Q} = \mathbf{G}^T (\mathbf{G} \mathbf{G}^T)^{-1} \mathbf{G}$ is the projector

onto the image space of \mathbf{G}^T (i.e. $\text{Im } \mathbf{Q} = \text{Im } \mathbf{G}^T$ and $\text{Im } \mathbf{P} = \text{Ker } \mathbf{G}$). We call the action of $(\mathbf{G}\mathbf{G}^T)^{-1}$ the *coarse problem* (CP) of FETI. It holds that $\mathbf{P}\boldsymbol{\mu} = \boldsymbol{\mu}$ because $\boldsymbol{\mu} \in \text{Ker } \mathbf{G}$, so the final problem reads

$$\mathbf{P}\mathbf{F}\boldsymbol{\mu} = \mathbf{P}(\mathbf{d} - \mathbf{F}\tilde{\boldsymbol{\lambda}}). \quad (4)$$

This final problem can be solved with an arbitrary iterative linear system solver. The conjugate gradient method is a good choice thanks to the classical estimate by Farhat, Mandel and Roux of the spectral condition number:

$$\kappa(\mathbf{P}\mathbf{F}\mathbf{P}|\text{Im}\mathbf{P}) \leq C \frac{H}{h}, \quad (5)$$

where h is the discretization parameter and H is the decomposition parameter. This estimate remains valid for TFETI.

3 The coarse problem solution strategies

Parallelization is achieved mainly by distributing diagonal blocks of \mathbf{K} over processors – each block reflects a subdomain. We strive to maximize the number of subdomains to reduce the sizes of the subdomain stiffness matrices, accelerating their factorization and \mathbf{K}^\dagger actions. Furthermore, thanks to the estimate (5), decomposition into more subdomains maintaining the fixed discretization leads to reduction of the condition number \mathbf{K} and thus the number of iterations.

A drawback is increasing null space dimension decelerating the CP solution $\mathbf{G}\mathbf{G}^T \mathbf{x} = \mathbf{y}$. It is kind of a communicating vessels effect. Furthermore, for a sufficiently large number of subdomains, the CP matrix $(\mathbf{G}\mathbf{G}^T)$ may be too large to fit into the memory of one computational unit. Unfortunately, the CP matrix is not block-diagonal. Thus, it is inevitable to use some *parallel* sparse direct solver for the CP solution. We have suggested and compared several strategies for parallel CP solution in [7]:

1. directly using LU or Cholesky factorization,
2. applying explicit inverse of $\mathbf{G}\mathbf{G}^T$,
3. iteratively using conjugate gradients (with Jacobi preconditioner),
4. orthonormalizing rows of G so that the CP is eliminated.

As noted above, the iterative approach is not suitable because it destroys robustness of the FETI method. The orthonormalization approach with the classical Gram-Schmidt method starts to fail when the nullspace is large enough (thousands) because round-off errors become an issue whereas the modified or iterative Gram-Schmidt methods have better numerical properties but are less scalable. So we have abandoned this approach so far. Therefore, we will further speak only about the first two strategies.

Concerning use of a direct solver, the CP dimension is not large enough to justify the fully parallel approach, i.e. using the whole global communicator – communication would take over computation in this case. Instead, we propose a

proper *partial* parallelization of this CP solution, i.e. using groups of processes (subcommunicators). We divide all processes of the global `PETSC_COMM_WORLD` communicator into the subcommunicators using PETSc built-in "pseudopreconditioner" `PCREDUNDANT`; the number of these subcommunicators is N_r (number of cores doing redundant work); this means the number of cores in each subcommunicator is $\approx N_c/N_r$. Let us now describe the two strategies in detail.

Strategy 1. \mathbf{GG}^T is factorized in the preprocessing phase. During the solution phase, each application of $(\mathbf{GG}^T)^{-1}$ consists of the forward and backward substitution using a parallel direct solver:

$$(\mathbf{GG}^T)^{-1}\mathbf{w} = \mathbf{U}^{\mathbf{GG}^T} \setminus (\mathbf{L}^{\mathbf{GG}^T} \setminus \mathbf{w}).$$

Strategy 2. A parallel direct solver is employed for the computation of the *explicit inverse* of \mathbf{GG}^T . During the preprocessing phase, \mathbf{GG}^T is factorized and then the explicit inverse $(\mathbf{GG}^T)^{-1}$ is computed. In the solution phase, its application consists in the parallel dense matrix-vector product $(\mathbf{GG}^T)^{-1}\mathbf{w}$.

The explicit inverse is assembled in the following way. Each of N_r subcommunicators is assigned a contiguous portion of N_n/N_r columns of the identity matrix taken as RHS. The result of the forward/backward substitutions is the corresponding portion of N_n/N_r columns of the resulting explicit inverse $(\mathbf{GG}^T)^{-1}$, stored as a $N_n \times (N_n/N_r)$ dense matrix distributed vertically across the subcommunicator. Taking advantage of the symmetry of $(\mathbf{GG}^T)^{-1}$, each subcommunicator's block is transposed in parallel and the blocks are then merged one below the other in the proper order forming the complete $(\mathbf{GG}^T)^{-1}$ matrix, distributed vertically across the global communicator.

4 Numerical experiments

We compare here both CP solution strategies using MUMPS and SuperLU direct solvers. The experiments were realized with our FLLOP (FETI Light Layer on top of PETSc) library[17] on HECToR supercomputer[16]. Matrices and vectors for numerical experiments were obtained from a regular decomposition and discretization of a model problem of an elastic cube. We discretized the cube into 4,096,000 elements and decomposed into 8,000 subdomains. The regular discretization and decomposition was chosen to ensure an equal workload of all cores. As it was already mentioned, significant efficiency improvement can be achieved by means of the partial parallelization of the CP solution. The performance results of parallel direct solvers for the decomposition into 8,000 subdomains and varying sizes of subcommunicators are shown in Table 1, 1 and Figure 4, 4 (log. scale). The optimal number of cores per subcommunicator for our model problem is 800, corresponding to $N_r=10$.

5 Conclusion

Without CP parallelization we are not able to solve large problems because of memory limitations and lack of scalability. So there is no other way than using

N_r		2000	1000	500	125	20	10	5
N_c/N_r		4	8	16	64	400	800	1600
fact.	M	23.3	15.4	16.1	17.7	11.8	12.2	9.1
	S	29.4	20.4	16.2	9.1	6.4	6.6	6.8
action	M	0.90	0.41	0.26	0.20	0.14	0.18	0.23
	S	1.03	0.50	0.41	0.28	0.14	0.06	0.11
+100 act.	M	114	56.1	42.1	37.9	25.4	29.7	32.5
	S	132	70.3	56.9	36.7	20.8	12.7	17.3
+1000 act.	M	926	422	276	220	148	187	243
	S	1059	519	423	285	150	68.2	112

N_r		2000	1000	500	125	20	10	5
N_c/N_r		4	8	16	64	400	800	1600
N_n/N_r		24	48	96	384	2400	4800	9600
fact.	M	23.3	15.4	16.1	17.7	11.8	12.2	9.1
	S	29.4	20.4	16.2	9.1	6.4	6.6	6.8
expl. inv.	M	21.7	19.5	25.0	77.6	326	840	2246
	S	24.7	24.0	39.1	106	346	296	1008
redist.	M	0.02	0.03	0.04	0.11	2.1	2.5	3.5
	S							
action		0.0072						
prep.	M	45.7	35.7	41.8	96.1	341	855	2260
	S	54.9	45.1	56.0	116	355	305	1019
+100 act.	M	52.2	42.2	48.3	103	348	862	2266
	S	61.4	51.6	62.6	122	361	312	1026

Table 1. Performance of MUMPS (M) and SuperLU_DIST (S) for Strategy 1 (upper) and Strategy 2 (lower) depending on the subcommunicator’s size for the decomposition into 8,000 subdomains (in seconds); the best variant is printed in bold.

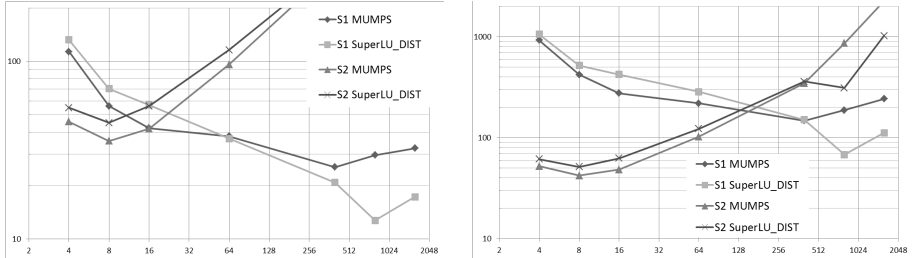


Fig. 1. Times for CP preprocessing + 100 $(\mathbf{G}\mathbf{G}^T)^{-1}$ actions (left) and CP preprocessing + 1000 $(\mathbf{G}\mathbf{G}^T)^{-1}$ actions (right) depending on the subcommunicator’s size, strategy, and direct solver (decomposition into 8,000 subdomains).

some parallel solver. On the other hand, engaging all processes in “world” communicator into the CP solution leads to an enormous communication overhead. Therefore, we conclude that the CP should be solved only in the subcommunicators of appropriate size. Two strategies for CP solution were introduced. The choice depends on the expected number of iterations given by the class of the solved problem. For more than 500 iterations, Strategy 2 with more expensive preprocessing starts to pay off. For the used Cray XE6 architecture and our PETSc-based implementation (FLLOP [17]) we can recommend following approaches: Strategy 1: SuperLU_DIST with $N_r=16$; Strategy 2: MUMPS with $N_r = 1000$.

References

1. T. Kozubek, V. Vondrák, M. Menšík, D. Horák, Z. Dostál, V. Hapla, P. Kabelíková, M. Čermák: “Total FETI domain decomposition method and its massively parallel implementation”, in *Advances in Engineering Software*, 2013.
DOI: 10.1016/j.advengsoft.2013.04.001
2. V. Hapla, D. Horak: “A Comparison of FETI Natural Coarse Space Projector Implementation Strategies”, in proceedings of PARENG2013, Paper 6, 2013.
DOI: 10.4203/ccp.101.6
3. D. Horak, V. Hapla: “A Comparison of Different Parallel Solvers for Linear Systems of Equations based on Domain Decomposition”, in proceedings of PARENG2013, Paper 3, 2013. DOI: 10.4203/ccp.101.3
4. V. Hapla, D. Horak, M. Merta: “Use of direct solvers in TFETI massively parallel implementation”, in proceedings of PARA 2012, Lecture Notes in Computer Science, 7782 LNCS, pp. 192–205, 2013. DOI: 10.1007/978-3-642-36803-5_14
5. M. Merta, A. Vasatova, V. Hapla, D. Horak: “Massively parallel implementation of Total-FETI method DDM with application to medical image registration”, in proceedings of DD21 2012, Lecture Notes in Computational Science and Engineering, accepted 2013.
6. D. Horák, V. Hapla: TFETI coarse problem massively parallel implementation, ECCOMAS 2012 e-Book Full Papers, pp. 8260–8267, ISBN: 978-395035370-9, 2012.
7. V. Hapla, D. Horák: “TFETI coarse space projectors parallelization strategies” in proceedings of PPAM 2011, Lecture Notes in Computer Science, 7203 LNCS (PART 1), pp. 152–162, 2012, DOI: 10.1007/978-3-642-31464-3_16
8. Z. Dostál: “Optimal Quadratic Programming Algorithms, with Applications to Variational Inequalities”, 1st edition. SOIA 23. Springer US, New York, 2009.
9. Z. Dostál, T. Kozubek: “An optimal algorithm and superrelaxation for minimization of a quadratic function subject to separable convex constraints with applications”, *Mathematical Programming*, 2011.
10. Z. Dostál, D. Horák, R. Kučera: “Total FETI – an easier implementable variant of the FETI method for numerical solution of elliptic PDE”, *Commun. in Numerical Methods in Engineering* 22, pp. 1155–1162, 2006. DOI:10.1002/cnm.881
11. R. Kučera, T. Kozubek, A. Markopoulos, On large-scale generalized inverses in solving two-by-two block linear systems, *Linear Algebra and its Applications*, Volume 438, Issue 7, 1 April 2013, pp. 3011–3029, ISSN 0024-3795, 2013. DOI:10.1016/j.laa.2012.09.027.
12. T. Brzobohatý, Z. Dostál, T. Kozubek, P. Kovář, A. Markopoulos: “Cholesky decomposition with fixing nodes to stable computation of a generalized inverse of the stiffness matrix of a floating structure”, *International Journal for Numerical Methods in Engineering*, Volume: 88, Issue: 5, pp. 493–509, 2011. DOI: 10.1002/nme.3187
13. Z. Dostál, T. Kozubek, A. Markopoulos, M. Menšík: “Cholesky decomposition of a positive semidefinite matrix with known kernel”, *Applied Mathematics and Computation*, Volume: 217, Issue: 13, pp. 6067–6077, 2011. DOI: 10.1016/j.amc.2010.12.069
14. S. Balay, J. Brown, K. Buschelman, V. Eijkhout, W.D. Gropp, D. Kaushik, M.G. Knepley, L.C. McInnes, B.F. Smith, and H. Zhang: “PETSc users manual”. Tech. Rep. ANL-95/11 - Revision 3.2, Argonne National Laboratory, 2011.
15. “PETSc Web page”, <http://www.mcs.anl.gov/petsc/>
16. “HECToR Web Page”, <http://www.hector.ac.uk/>
17. “FLLOP Web Page”, <http://spomech.vsb.cz/feti/>

Comparison of laparoscopic and open surgery techniques in liver surgery

Martin Hasal

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
`martin.hasal@vsb.cz`

Abstract. Statistical comparison of laparoscopic and open surgery techniques used in Radiofrequency-assisted liver resections is an important task to assess which techniques are more onerous for patient. Data was collected during five years and they were processed according to the requirements of doctors according to the global parameters, which are mentioned in literature. The paper presents exploratory analysis of data, comparison of parameters for both approaches to assess if laparoscopy approach is not more onerous performance than open approach in liver resections. Kaplan-Meier statistics for survival is used as parameter of learning of surgeon during the two periods of study.

Key words: Statistical comparison, liver resection, open approach, laparoscopic approach, Kaplan-Meier statistics, disease-free interval.

1 Introduction

Experimental use of radiofrequency (RF) ablation in liver tumours was pioneered using of RF to treat primary liver tumours, when RF was used for ablation of intra parenchymatous liver lesions with minimal harm to the surrounding area. Based on this success, RF ablation using USG was began to use as oncological therapy for tumours in other organs, such as prostate tumours, lung tumours, bone tumours or brain metastases [1], [2], [3]. And after it RF have also been used for RF-assisted liver resection ('Habib's resection'). [4], [5]

Firstly RF-assisted liver resection was done using an open approach of surgery. And even after twenty years liver surgery still remains one of the last areas of resistance to the offensive of laparoscopy. RF-assisted laparoscopic liver resection has been recently developed technique for treatment of primary and secondary liver tumors. Nowadays minimally invasive approaches are widely used in other surgical areas, at University Hospital Ostrava it was decided, that it is time to assess the effectiveness of using of laparoscopy approach also in RF-assisted liver resection. Further only the names of laparoscopic liver resection (LLR) for laparoscopic RF-assisted liver resection and open liver resection (OLR) for open RF-assisted liver resection are used [5], [6], [7], [8].

2 Materials and methods

It was done a systematic prospective study at University Hospital Ostrava from January 2010 to December 2012. Data were stored using a MySQL database (Oracle, Redwood Shores, CA, USA) and analysed using SPSS version 18 (IBM, Rochester, MN, USA). For providing basic characteristics of the sample descriptive statistics was performed. The goodness-of-fit test (χ^2) was used to explore whether the data were normally distributed. For testing the difference between the groups the Mann-Whitney test was used (normal distribution was not demonstrated). Kaplan-Meier statistic (Log rank, Breslow and Tarone-Ware) were performed to test differences in survival and disease-free interval. For all tests, the significance level was chosen as 0.05 [9].

The null hypothesis says that laparoscopic liver resections and open liver resections do not differ in medians of operative parameters, such as blood loss, operative time, length of stay at the hospital and at the intensive care unit and in survival and disease-free interval. On the other hand the alternative hypothesis says that laparoscopic liver resections show better outcomes in medians of operative parameters, survival and disease-free than open liver surgery [10].

3 Results

During the period, 100 patients underwent liver surgery. 26 patients of them were excluded from the study because of other intervention on the liver such as marsupialization of liver cyst, application of gold seeds for CyberKnife therapy or liver biopsy [5]. Thus, 74 patients with primary or secondary liver cancers and symptomatic benign liver lesions were enrolled in the study. These patients was divided into two groups - 54 patients underwent liver resection through open surgery and 20 patients underwent laparoscopic surgery. The number of patients according to the operative approach is shown in Figure 1 [10], [11].

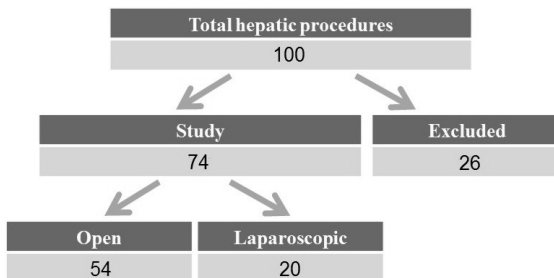


Fig. 1. The operative approaches used in the hepatic surgeries in this study.

In the study group of 74 patients, there were 38 men (51.4%) and 36 women (48.6%). The mean patient age for the laparoscopic approach was 57.5 years, and the mean patient age for the open approach was 60 years.

3.1 Operative and postoperative characteristics for laparoscopic vs. open approach

As a operative characteristics length of stay in the hospital (LSH), length of stay in the intensive care unit (LSICU), blood loss (BL), operative time (OT) and drain leaking 1. day and 2. day (DL) were chosen [5], [9].

As it was mentioned the main idea is that the laparoscopic liver resection is less onerous performance with better short-term postoperative results than open liver resection.

The hypothesis was separately selected for two groups of parameters. Because for parameter of LSH the medians was for both groups the same. For LSH, BL and OT the hypotheses say that

H_0 : Between the medians for laparoscopic and open surgery approach the statistical significant difference does not exist for the selected parameters, on the other hand

H_A : Median for laparoscopic surgery approach is lower than for open surgery approach and the difference is statistical significant.

The result of the tests are depicted in Table 3.1. From the results of the tests it can be said that the observed difference in medians for LSICU and BL is significant on chosen significant level. And for OT and DL the observed difference in medians is not significant.

Table 1. The results of Mann-Whitney test for assessment of difference in analysed parameters [11].

Analyzed parameter	p-value
Length of stay in the intensive care unit	0,042
Blood loss	0,037
Operative time	0,121
Drain leaking 1. day and 2. day	0,386

As the medians for open and laparoscopic approach for LSH are the same the hypotheses say that

H_0 : Between the medians for laparoscopic and open surgery approach the statistical significant difference does not exist for the LSH, on the other hand

H_A : Between the medians for laparoscopic and open surgery approach the statistical significant difference exists for the length of stay in the hospital.

The result of the test is determined as p-value=0.283. From the result of the test it can be said that the observed difference in medians for LSH is not significant on chosen significant level.

3.2 Kaplan-Meier survival analysis-learning curves

Survival of study patients were compared with survival of patients from previous period 2007-2009 using the Kaplan-Meier survival analysis. The assessment of difference is used as a assessment of learning of the surgeon. As during the time, the surgeon has more and more experiences and the result of operation represented by the survival could get better. The first period 2007-2009 is represented by 60 and the second period 2010-2012 by 74 patients. All patients had result of operation R0. Patients are mixed, laparoscopic approach and open approach together (Fig. 2).

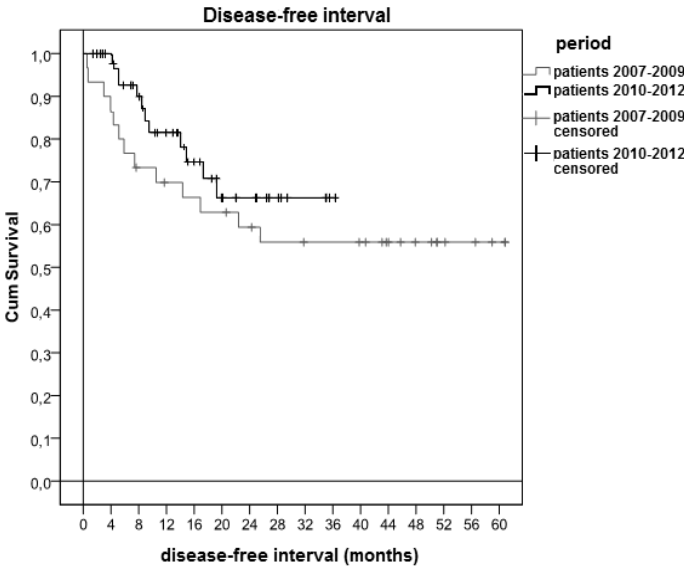


Fig. 2. The Kaplan-Meier survival curves for patients with metastases and R0 operation according to the operative approach [7].

Table 3.2 shows survival for both period. The mean survival time and 5-year survival is greater for the second period 2010-2012 than for the first period 2007-2009.

Table 2. The results of Mann-Whitney test for assessment of difference in analysed parameters.

Operative approach	Mean survival time (months)	5-year survival (%)	Test p-value		
Period 2007-2009	38.9	56.7	Log rank	Breslow	Tarone-Ware
Period 2010-2012	26.4	67.2			

The significance of the difference in survival time in groups were analysed using the log rank, Breslow, and Tarone-Ware tests; none of the tests showed the difference between groups to be significant (Table 3.2).

4 Conclusion

According to the results presented in the paper it could be said, that the laparoscopic approach for RF-assisted liver resection is not more onerous performance for patient, even in some operative and postoperative parameters laparoscopic approach shows better operative and short-term postoperative results than open liver resection. The better results are achieved in length of stay in the intensive care unit and blood loss. The reducing length of stay in intensive care unit could cause reducing the major part of costs and brings higher comfort for patient. It is part of overall length of stay, but intensive care is one of the most expensive of hospital care. The learning curves do not show any significant difference between groups formed by used period, but the 5-year survival for the second period 2010-2012 is longer by approximately 10 months, which is not negligible for patients.

Acknowledgements This work has been supported by Project SP2013/191, "Parallel solving computationally intensive tasks", of the Student Grant System, VŠB - Technical University of Ostrava.

References

1. Karasawa, K., Muta, N., Nakagawa, K., Hasezawa, K., Terahara, A., Onogi, Y et al. *Thermoradiotherapy in the treatment of locally advanced non small cell lung cancer*. International Journal Radiation Oncology Biology Physics 1994; 30: 1171-1177.
2. Anzai, Y., Lufkin, R., DeSalles, A., Hamilton, DR., Farahani, K., Black, KL. *Preliminary experience with MR guided thermal ablation of brain tumors*. American Journal of Neuroradiology 1995; 16: 39-48.
3. Gangi, A. *The treatment of osteoid osteoma: a multitude of choice: surgery, percutaneous resection, alcohol injection or thermocoagulation*. Journal of Radiology 1999; 80: 419-420.
4. Pai, M., Frampton, AE., Mikhail, S., Resende, V., Kornasiewicz, O., Spalding, DR. et al. *Radiofrequency assisted liver resection: Analysis of 604 consecutive cases*. European Journal of Surgical Oncology 2012; 38: 274-280.
5. Vávra, P., Nowaková, J., Jelínek, P., Hasal, M., Penhaker, M., Ihnát, P., Jurčíková, J., Habib, N., Zonča, P. *Radiofrequency/assisted liver resections: comparison of open and laparoscopic techniques*. Journal of Laparoendoscopic & Advanced Surgical Techniques - in review.
6. Vávra, P., Twardziková, P., Penhaker, M., Vávrová, M., Jonszta, T., Hasal, M., Nowaková, J., Ihnát, P., Habib, N. *Hepatocellular Carcinoma: Options and Limits*. VI. SCHK a XXXIX. Spoločný zjazd ČS chirurgov s medzinárodnou účasťou, 12.-14.9.2012, Košice, Slovensko. Supplement p.35, ISBN 978-80-7097-969-3.

7. Vávra, P., Hasal, M., Ihnát, P., Nowaková, J., Czerný, D., Jelínek, P., Penhaker, M., Ostruszka, P., Habib, N. *Laparoscopic Liver Resections*. VI. SCHK a XXXIX. Společný zjazd ČS chirurgov s medzinárodnou účasťou,, 12.-14.9.2012, Košice, Slovensko. Supplement p.4, ISBN 978-80-7097-969-3.
8. Vávra, P., Twardziková, P., El-Gendi, A., Vávrová, M., Nowaková, J., Jonszta, T., Hasal, M., Ihnát, P., Habib, N. *Možnosti léčby hepatocelulárního karcinomu*. 12. Česko Polsko Slovenský kongres miniinvazivní chirurgie, 9. Ostravské dny miniinvazivní chirurgie, 1.11.2012 - 02.11.2012, Ostrava.
9. Cipryan L., Litschmannová, M. *Intra-day and inter-day reliability of heart rate variability measurement*. Journal of Sports Science and Medicine 2013; 31: 150-8.
10. Vávra, P., Nowaková, J., Hasal, M., Jonszta, T., Vávrová, M., Ihnát, P., Penhaker, M., Habib, N. *Srovnání otevřené a laparoskopické radiofrekvenčně asistované resekce jater*. XXIV. Petřivalského-Rapantův den, 26.-27.4.2012, Olomouc. Supplement p.16.
11. Vávra, P., Nowaková, J., Hasal, M., Penhaker, M., Ihnát, P., Ostruszka P., Jelínek P. *Resekce jaterních tumorů: srovnání laparoskopického a otevřeného přístupu*. 12. Česko Polsko Slovenský kongres miniinvazivní chirurgie, 9. Ostravské dny miniinvazivní chirurgie, 1.11.2012 - 02.11.2012, Ostrava. Supplement p.28.

Identification problems with a priori given material interfaces

Rostislav Hrtus^{1,2}, Jaroslav Haslinger¹, and Radim Blaheta¹

¹Institute of Geonics AS CR, Department IT4Innovations

²Department of Applied Mathematics, FEECS,

VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{hrtus, blaheta}@ugn.cas.cz

Abstract. The paper investigates material parameter identification problems based on elliptic state problem with a priori known material interfaces. This investigation covers both scalar and elastic state problems and distributed type cost functions. A class of important applications, where such identification problems appear, is shown to be identification of local material properties of (geo)composites with material interfaces given a priori from tomography scans.

1 Preliminary

Our aim is to describe and investigate material parameter identification for elliptic state problems with a priori known material interfaces. For both scalar elliptic problem and elasticity, we describe the choice of set of admissible parameters with favourable properties due to restriction to the case of a priori known material interfaces. We formulate identification problems using the above state problems, definitions of sets of admissible parameters and distribution type criteria of fitting the state problem solution with experimental data. Such identification problems are shown to be solvable. Further, we shall also consider the finite element approximation of the state problem and investigate the mutual relation between approximate and exact identification problems.

Finally, we describe a problem of identification of local material properties of (geo)composite, see [3]. This problem has material interfaces given a priori from tomography scans and use the distributed data fit criterion. Therefore, it fits both the class of investigated identification problems and provides an example of important engineering applications.

2 Identification problem with given material interfaces

This section gives a description of parameter identification problem in the case of a priori given material interfaces and summarize theoretical results from [4].

We consider a **state problem**

$$\begin{cases} -\operatorname{div}[a(\mathbf{x})\nabla u(\mathbf{x})] = f(\mathbf{x}) & \text{for } \mathbf{x} \in \Omega \subset \mathbb{R}^d, \\ u(\mathbf{x}) = 0 & \text{on } \partial\Omega, \end{cases} \quad (1)$$

where $d = 1, 2, 3$, $f \in L^2(\Omega)$ and the coefficients $a = a(\mathbf{x}) \in \mathcal{Q}$, where

$$\mathcal{Q} = \{a \in L^\infty(\Omega) \mid a_{min} \leq a(\mathbf{x}) \leq a_{max}, \forall \mathbf{x} \in \Omega\}. \tag{2}$$

Here a_{min} and a_{max} are given positive values such that $0 < a_{min} < a_{max}$. The homogenous Dirichlet boundary condition is used for simplicity and both the identification problem formulation and theoretical results can be easily generalized for more general (mixed) boundary conditions.

The weak formulation of (1) is given in the following way

$$\mathcal{P}(a) \quad \begin{cases} \text{find } u(a) \in H_0^1(\Omega) \text{ s.t.} \\ \int_{\Omega} a \nabla u(a) \nabla \phi \, dx = \int_{\Omega} f \phi \, dx \quad \forall \phi \in H_0^1(\Omega), \end{cases}$$

It is important that for any $a \in \mathcal{Q}$ there is a unique solution $u(a)$ of $\mathcal{P}(a)$.

To define the set of **admissible control variables**, we assume **a priori given material interfaces**, i.e. availability of a disjunctive decomposition $\Omega = \bigcup_{i=1}^r \bar{\Omega}_i$, $\Omega_i \cap \Omega_j = \emptyset$ for $i \neq j$, such that the control variable a is constant in Ω_i .

Therefore,

$$\mathcal{Q}_{ad} = \{a \in \mathcal{Q} \mid a|_{\Omega_i} = a^i \in \mathbb{R} \quad \forall i = 1, \dots, r\}, \tag{3}$$

Thus $a \in \mathcal{Q}_{ad}$ can be represented by an algebraic vector, $a = (a_1, \dots, a_r) \in \mathbb{R}^r$. It is further important that for $a_n, a \in \mathcal{Q}_{ad}$, $a_n \equiv (a_{n_1}, \dots, a_{n_r})$, $a \equiv (a_1, \dots, a_r)$, we introduce a componentwise convergence, which will be equivalent to L^∞ convergence,

$$a_n \rightarrow a \stackrel{\text{df}}{\Leftrightarrow} a_{n_i} \rightarrow a_i \quad \forall i = 1, \dots, r \Leftrightarrow \|a_n - a\|_{L^\infty} \rightarrow 0.$$

As a direct consequence of Bolzano-Weierstrass theorem, \mathcal{Q}_{ad} is a sequentially compact set.

Identification problems also use **data-fit criterion or a cost function** which is usually a least squares type functional, which measures the discrepancy between model output (based on the solution of $\mathcal{P}(a)$) and data coming from observations or measurements of the real (physical) state. We shall consider here the case of distributed criteria, which means that the observation data have the form of functions on Ω and the cost functions have e.g. the form

$$\mathcal{J}_0(a) = \frac{1}{2} \left\| u(\mathbf{x}, a) - z^{(0)}(\mathbf{x}) \right\|_{2,0}^2 \quad \text{or} \quad \mathcal{J}_1(a) = \frac{1}{2} \left\| \nabla u(\mathbf{x}, a) - z^{(1)}(\mathbf{x}) \right\|_{2,0}^2 \tag{4}$$

where $a \in \mathcal{Q}_{ad}$, $u(a)$ it the solution of $\mathcal{P}(a)$, $z^{(0)}$ and $z^{(1)}$ are given functions derived from the measurements, $z^{(i)} \in L_2(\Omega)$.

Now we are ready to formulate the **identification problem (IP)**

$$(\mathcal{IP})_i \quad \begin{cases} \text{find } a \in \mathcal{Q}_{ad} \text{ s.t.} \\ \mathcal{J}_i(a) \leq \mathcal{J}_i(\tilde{a}) \quad \forall \tilde{a} \in \mathcal{Q}_{ad} \end{cases}$$

Note that compactness of \mathcal{Q}_{ad} and continuity of the mapping $P : a \rightarrow u(a)$, $P : \mathcal{Q}_{ad} \rightarrow H_0^1(\Omega)$ and the mapping $\mathcal{J}_i : a \rightarrow \mathcal{J}_i(a)$, $\mathcal{J}_i : \mathcal{Q}_{ad} \rightarrow \mathbb{R}^1$ ($i = 1, 2$) allows to prove the following theorem.

Theorem 1. *For $i = 0, 1$, problem $(\mathcal{IP})_i$ has a solution $a \in \mathcal{Q}_{ad}$.*

Proof. See [4].

Practically we solve $\mathcal{P}(a)$ only approximately, frequently by using the finite element method. Thus we can consider a sequence of regular triangulations $\{\mathcal{T}_{h_j}\}$, $h_j \rightarrow 0$, such that the elements of \mathcal{T}_{h_j} do not cross the material interfaces. If $V_{h_j} \subset H_0^1(\Omega)$ denote the corresponding finite element spaces, then $\mathcal{P}(a)$ can be replaced by

$$\mathcal{P}(a)_h \quad \begin{cases} \text{find } u_h \in V_h \text{ s.t.} \\ \int_{\Omega} a \nabla u_h \nabla \phi \, dx = \int_{\Omega} f \phi \, dx \quad \forall \phi \in V_h, \end{cases}$$

and the identification problems take the form

$$(\mathcal{IP}_h)_i \quad \begin{cases} \text{find } a_h \in \mathcal{Q}_{ad} \text{ s.t.} \\ \mathcal{J}_{i_h}(a_h) \leq \mathcal{J}_{i_h}(\tilde{a}_h) \quad \forall \tilde{a}_h \in \mathcal{Q}_{ad}, \end{cases}$$

where $\mathcal{J}_{0_h}(a) = \frac{1}{2} \|u_h(\mathbf{x}, a) - z^{(0)}(\mathbf{x})\|_{2,0}^2$, $\mathcal{J}_{1_h}(a) = \frac{1}{2} \|\nabla u_h(\mathbf{x}, a) - z^{(1)}(\mathbf{x})\|_{2,0}^2$, $u_h(a)$ is the solution of $\mathcal{P}(a)_h$.

Theorem 2. *let $\{a_{n_j}\}$ be a sequence of solutions of (\mathcal{IP}_{h_j}) . Then there is a subsequence $\{a_{h_{j_k}}\}$ such that for $k \rightarrow \infty$:*

$$\begin{aligned} a_{h_{j_k}} &\rightarrow a^*, \quad a^* \in \mathcal{Q}_{ad}, \\ u_{h_{j_k}}(a_{h_{j_k}}) &\rightarrow u(a^*) \text{ in } H_0^1(\Omega). \end{aligned} \tag{5}$$

Moreover, $u(a^*)$ is the solution of $\mathcal{P}(a^*)$ and a^* solves (\mathcal{IP}) .

Proof. See [4].

Note that beside the discussed topics, we are also interested in the numerical realisation of the approximate identification problems (\mathcal{IP}_h) , see e.g. [2], as well as stability of the solution (sensitivity to the parameters) and possible regularization, see e.g. [1].

3 Identification problem in elasticity

Now, we shall consider **elasticity state problem**, which can be written in the following form

$$\begin{aligned} -\operatorname{div}[A(\mathbf{x})e(u(\mathbf{x}))] &= f(\mathbf{x}) \text{ for } \mathbf{x} \in \Omega \subset \mathbb{R}^d, \\ u(\mathbf{x}) &= 0 \text{ on } \partial\Omega, \end{aligned} \tag{6}$$

or variationally as

$$\mathcal{P}_e(A) \quad \begin{cases} \text{find } u \in H_0^1(\Omega)^d \text{ s.t.} \\ \int_{\Omega} \langle A(\mathbf{x})e(u(A)), e(\phi) \rangle dx = \int_{\Omega} \langle f, v \rangle dx \quad \forall v \in H_0^1(\Omega)^d. \end{cases}$$

Above, $u : \Omega \rightarrow \mathbb{R}^d$ denotes the displacement, $f : \Omega \rightarrow \mathbb{R}^d$ represents density of volume forces, $A(\mathbf{x})$ is the elasticity tensor, $\langle \cdot, \cdot \rangle$ denotes the euclidean inner product in \mathbb{R}^d . We assume $A \in \mathcal{Q}^e$,

$$\mathcal{Q}^e = \{A = [a_{ij}(\mathbf{x})], a_{ij} = a_{ji} \in L^\infty(\Omega), \nu \leq \lambda_{\min}(A), |a_{ij}| \leq \varkappa \forall ij\}, \tag{7}$$

where ν, \varkappa are given positive constants, $\lambda_{\min}(A)$ denotes the smallest eigenvalue of A .

Note that $\mathcal{P}_e(A)$ has exactly one solution for any $A \in \mathcal{Q}^e$, see e.g. [5]. Again, the homogenous Dirichlet boundary condition is used for simplicity and generalization to other types of boundary conditions is possible.

To define the set of **admissible control variables**, we assume again a priori given material boundaries, i.e. a disjunctive decomposition $\{\Omega_i\}_{i=1}^r$ of Ω such that the control variable is constant in Ω_i . Correspondingly,

$$\mathcal{Q}_{ad}^e = \{A \in \mathcal{Q} \mid A|_{\Omega_k} = A_k = [a_{ij}^k] \in \mathbb{R}^{d \times d} \quad \forall k = 1, \dots, r\}, \tag{8}$$

Thus $A \in \mathcal{Q}_{ad}^e$ can be represented as an array of symmetric matrices

$$A \equiv [a_{ij}^k] \in \mathbb{R}^{r \times d \times d}, a_{ij}^k = a_{ji}^k. \tag{9}$$

Let us define a componentwise convergence on \mathcal{Q}_{ad}^e , i.e. for $A_n, A \in \mathcal{Q}_{ad}^e$, $A_n \equiv [a_{ij}^k]_n \equiv a_{ij,n}^k, A \equiv [a_{ij}^k]$,

$$A_n \rightarrow A \stackrel{\text{df}}{\iff} a_{ij,n}^k \rightarrow a_{ij}^k \quad \forall i, j = 1, \dots, d, k = 1, \dots, r. \tag{10}$$

Due to this definition

$$A_n \rightarrow A \iff \|A_n - A\|_{L^\infty(\Omega)} = \operatorname{ess\,supp} \|A_n(\mathbf{x}) - A(\mathbf{x})\|_{sp} \rightarrow 0, \tag{11}$$

where $\|M\|_{sp}$ is the spectral (euclidean) norm of a matrix M .

It is important that \mathcal{Q}_{ad}^e is closed. If $A_n, A \in \mathcal{Q}_{ad}^e, A_n \rightarrow A$, then

$$|a_{ij,n}| \leq \kappa \Rightarrow |a_{ij}| \leq \kappa, \tag{12}$$

$$\lambda_{min}(A_n) \geq \nu \Rightarrow \lambda_{min}(A) \geq \nu, \tag{13}$$

i.e. $A \in \mathcal{Q}_{ad}^e$. The last implication is a consequence of a continuous dependence of eigenvalues on the matrix entries, see e.g. [6]. Due to componentwise character of the convergence, \mathcal{Q}_{ad}^e is again a sequentially compact set.

The **cost function** can be defined similarly to (4). The simplest choice is

$$\mathcal{J}_0^e(A) = \frac{1}{2} \left\| u(A, \mathbf{x}) - z^{(e)} \right\|_{L_2(\Omega)^d}^2. \tag{14}$$

Another choices are e.g.

$$\mathcal{J}_2^e(A) = \frac{1}{2} \|e(A, \mathbf{x}) - \hat{e}\|_{L_2(\Omega)^{d \times d}}^2 \quad \text{and} \quad \mathcal{J}_3^e(A) = \frac{1}{2} \|\sigma(A, \mathbf{x}) - \hat{\sigma}\|_{L_2(\Omega)^{d \times d}}^2, \tag{15}$$

where $\|v\|_{L_2(\Omega)^d}^2 = \int_{\Omega} \|v\|^2 dx, \|A\|_{L_2(\Omega)^{d \times d}}^2 = \int_{\Omega} \|A\|_{sp}^2 dx, \hat{e}$ and $\hat{\sigma}$ are measured strains and stresses.

The **identification problem** (\mathcal{IP}) can be now defined similarly as in Section 2, i.e. for $i = 0, 2, 3$

$$(\mathcal{IP})_i \quad \begin{cases} \text{find } a \in \mathcal{Q}_{ad}^e \text{ s.t.} \\ \mathcal{J}_i(a) \leq \mathcal{J}_i(\tilde{a}) \quad \forall \tilde{a} \in \mathcal{Q}_{ad}^e. \end{cases}$$

Note that compactness of \mathcal{Q}_{ad}^e and continuity of the mapping $P : a \rightarrow u(a), P : \mathcal{Q}_{ad}^e \rightarrow H_0^1(\Omega)$ and the mapping $\mathcal{J}_i : a \rightarrow \mathcal{J}_i(a), \mathcal{J}_i : \mathcal{Q}_{ad}^e \rightarrow \mathbb{R}^1 (i = 1, 2)$ allows to prove the existence of solution of the identification problem $(\mathcal{IP})_i$.

4 Identification of local material properties of (geo)composites

As an example of applications of the class identification problems with a priori known material interfaces, we show analysis of local material properties of (geo)composite. A coal - polyurethane geocomposite arising from grouting is tested for analysis of its effective (homogenized) mechanical properties. A $7 \times 7 \times 7$ cm sample is scanned by computer tomography (CT) and this scan is segmented to obtain the material interfaces, see [3]. An overall response can be then computed if we know the local material properties. These properties can be obtained by separate testing samples of pure materials of the constituents or using indentation techniques, but the obtained knowledge is far to be perfect. Therefore the use of parameter identification can be very useful.

If laboratory tests of some samples of heterogeneous composite material are available, then the identification can use cost function measuring the difference of experimentally obtained homogenized (constant) strains \hat{e} and stresses $\hat{\sigma}$ in

comparison with averaged micro strains $\langle e(u(A)) \rangle$ and micro stresses $\langle \sigma(u(A)) \rangle$ obtained by solving an elasticity state problem. The averaging operator is denoted by brackets, $\langle \eta \rangle = |\Omega|^{-1} \int_{\Omega} \eta(x) dx$. A suitable cost function has the form

$$\mathcal{J}_4^e(A) = \frac{1}{2} \|\langle e(u(A)) \rangle - \hat{e}\|_{L_2(\Omega)^{d \times d}}^2 + \frac{1}{2} \|\langle \sigma(u(A)) \rangle - \hat{\sigma}\|_{L_2(\Omega)^{d \times d}}^2$$

which provides again cost function depending continuously on the control variable $A \in \mathcal{Q}_{ad}^e$. Therefore the identification problem fits the class described in the previous sections.

Note that the approximate problem uses voxel grid corresponding to the CT scan and approximate now both the state problem solution and in some sense also the material interfaces. The discretized elasticity problem has millions of degrees of freedom, which indicates the importance of efficient numerical methods used for both the state problem solution and solution of the optimization problem, see e.g. [2] for more details.

5 Conclusions

This paper describes a class of identification problems with a priori given material interfaces and shows both theory and important applications of this class of identification problems. A future work will be directed to some generalization of this theory and topics of numerical solution, stability (sensitivity) and possible regularization. Note that some of our previous papers (see e.g. [2]) are devoted to the numerical solution including aspects of parallel computations.

Acknowledgement

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and by the project SPO-MECH - Creating a multidisciplinary R&D team for reliable solution of mechanical problems, reg. no. CZ.1.07/2.3.00/20.0070 within Operational Programme 'Education for competitiveness' funded by Structural Funds of the European Union and state budget of the Czech Republic.

References

1. H. T. Banks, K. Kunisch: Estimation Techniques for Distributed Parameter Systems. Birkhäuser, Boston-Basel-Berlin 1989, 315 pp.
2. Blaheta R., Hrtus, R., Kohut, R. Axelsson, O. Jakl, O. , Material parameter identification with parallel processing and geo-applications. In: R. Wyrzykowski et al. (Eds.): PPAM 2011, Part I, Springer-Verlag LNCS 7203, pp. 366–375
3. Blaheta R. et al., Micromechanics of geocomposites: CT images and FEM simulations. EUROCK 2013, to appear
4. Haslinger J.: Identification Methods. Manuscript (2013)
5. Necas J., Hlavacek I.: Mathematical Theory of Elastic and Elasto-Plastic Bodies: An Introduction. Elsevier, Amsterdam-Oxford -New York (1981)
6. Tyrtyshnikov, E.E. (1997). A Brief Introduction to Numerical Analysis. Birkhäuser Boston.

Weibull Proportional Hazards Regression Model for Patients with Surgery of Colon

Kateřina Janurova

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
katerina.janurova@vsb.cz

Abstract. Medical survival censored data of 876 patients are evaluated to compare two basic surgical techniques. Data comes from patients who underwent colectomy in the University Hospital in Ostrava. Two basic surgery techniques are used for the colectomy: either classical (open) or laparoscopic operation. Basic question which arises at the colectomy operation is which type of operation to choose to guarantee longer overall survival time. The parametric Weibull proportional hazards regression model has been used to answer this question, because of the fact that the hazard function of human life is often described as being “bathtub shaped”.

Keywords: survival analysis, right-censored medical survival data, Weibull proportional hazards regression model

1 Introduction

The goal of this article is the comparison of two basic surgical techniques, in order to answer the question: is there a type of surgery which guarantees longer overall survival time and which one is it?

The laparoscopic surgical technique has many advantages such as lower operative stress and more favorable post-operative course. On the other hand, there are significant disadvantages in using these techniques in colorectal surgery, which can participate in morbidity in large measure (e.g. the risk of capnoperitoneum, longer operative time and extreme positioning of patients). The results of various medical studies regarding the comparison of morbidity and mortality after both types of surgeries of colon are commonly available and proved the merits of laparoscopic surgical technique, e.g. [1], [2], [3], [4], [5] and [6], although the consensus of European association of endoscopic surgery for colon carcinoma mentions, that there is no difference between morbidity of laparoscopic and open operations of colon [7]. Regarding the surgeries of rectum, much less information is available nowadays. Meta-analyses comparing laparoscopic versus open surgery for rectal cancer are quite rare and in fact they are mostly connected with short term results. Therefore the analysis of the mortality after both types of surgeries of rectum is still open problem.

This article provides an analysis of medical survival data of 876 patients, who underwent surgical resection of colon or rectum at the University Hospital of Ostrava during the years 2001-2009 and were followed up until December of 2012. The data has been divided into two basic groups: patients operated by a classical (open) technique and patients operated by minimally invasive (laparoscopic) technique.

Data coming from the medical study, are right-censored, in the form of vectors $(t_i, c_i, x_i), \dots, (t_n, c_n, x_n)$, where t_i is either a time of death or a time in which the observation of i -th patient is stopped (withdrawn), $c_i = 1$ (resp. $c_i = 0$) is censoring indicator, according to death (resp. stopping time) occurring first and x_i is the indicator of the type of surgical technique, where $x_i = 0$ if the i -th patient was operated by the laparoscopic technique and $x_i = 1$ if the i -th patient was operated by the open technique. At censored observation the only available information is that the survival time of a patient is greater than observed value, under the assumption of non-informative censoring.

The surgical techniques were compared separately for patients with resection of colon and for patients with resection of rectum using the Weibull proportional hazards regression model.

2 Basic Relations in Survival Analysis

In applied survival analysis, we are typically interested in describing how long the patients live; therefore we concentrate on estimating of survival function. It represents the probability that the survival time of patient is greater than some specified time t , denoted $S(t)$. The distribution function, denoted as $F(t)$, on the other hand, represents the probability that the survival time is less than or equal to some value t , so

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) \quad (1)$$

The risk of death at some time t is expressed by the hazard function, denoted $h(t)$, which can be obtained from the probability that the patient dies at time t , under the condition of surviving to time t . The relationship between hazard and survival function is given by the formula

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} = \frac{-d \log S(t)}{dt} \quad (2)$$

where $f(t)$ is the probability density function, which can be expressed as $f(t) = F'(t) = -S'(t)$ for $t \geq 0$.

From Equation 2, the survival function can be obtained as

$$S(t) = \exp\left[-\int_0^t h(u) du\right] = \exp(-H(t)), \quad t \geq 0 \quad (3)$$

where $H(t)$ is called integrated or cumulative hazard function.

3 Weibull Proportional Hazards Regression Model

The Weibull regression model is the parametric version of the Cox proportional hazards model. The key difference between the models is that the survival time is assumed to follow the Weibull probability distributions when a fully parametric proportional hazards model is fitted to the data and that the parameters β are estimated by maximum likelihood method instead of the partial likelihood method.

If the survival time has the Weibull probability distribution with the shape parameter λ and scale parameter γ , then the baseline hazard function $h_0(t)$ can be written as $h_0(t) = \lambda \gamma t^{\lambda-1}$ and the hazard function for the Weibull regression model is given by

$$h(t|\mathbf{x}) = \lambda \gamma t^{\lambda-1} \exp(\mathbf{x}'\beta). \quad (4)$$

It can be seen from the form of the hazard function, given by Equation 4, that the survival time of the patient in the study has a Weibull distribution with the scale parameter $\gamma \exp(\mathbf{x}'\beta)$ and shape parameter λ . This shows that the scale parameter of the distribution is modified by the effects of the covariates; therefore the proportional hazards property is held, while the shape parameter remains constant [8].

The survival function corresponding to the proportional hazards form of the hazard function given by Equation 4 is found using Equation 3

$$S(t|\mathbf{x}, \beta) = \exp\{-\exp(\mathbf{x}'\beta)\gamma t^\lambda\}. \quad (5)$$

The equation for the median survival time can be obtained by setting the survival function equal to 0.5 and solving for time [9], yielding

$$t_{50}(t|\mathbf{x}) = \{\log(2)/[\gamma \exp(\mathbf{x}'\beta)]\}^{1/\lambda}. \quad (6)$$

Fitting the Weibull Proportional Hazards Regression Model is discussed in detail in [10] and the appropriateness of the Weibull assumption in [11].

4 Comparison of Survival Data of Patients with Surgery of Colon

Weibull proportional hazards model, given by Equation 4, containing single covariate was fitted to the survival data of 534 patients who underwent the surgery resection of colon. The results of model fitting are shown in Table 1.

Table 1. Estimated coefficients, standard error, z-score, two-tailed p-value for the Weibull regression model containing type of surgical technique for patients with resection of colon.

Variable	Value	Std. Err.	<i>z</i>	<i>p</i> > <i>z</i>
Type	-0.495	0.1477	-3.35	0.0008
Constant	8.184	0.1110	73.70	0.0000
ln(sigma)	0.251	0.0495	5.06	0.0000
sigma	1.285			

We may obtain the shape parameter λ as $1/\sigma$ [9], where parameter σ is a variance-like parameter on the log-scale, so $\lambda = 1/1.285 = 0.778$ and the scale parameter γ as $\exp(-\text{constant}/\sigma)$, so $\gamma = \exp(-8.184/1.285) = 0.0017$. The coefficient $\hat{\beta}$ is than a ratio of minus constant for type to sigma, what leads to $\hat{\beta} = 0.495/1.285 = 0.3852$.

Using results obtained above and Equation 4, we may write the hazard function for the Weibull regression model as

$$h(t|x) = 0.778 \cdot 0.0017 t^{0.778-1} \exp(0.3852 \cdot \text{type}) = 0.0013 t^{-0.222} \exp(0.3852 \cdot \text{type})$$

and using Equation 5, the survival function corresponding to the open type of surgical technique can be written as

$$S(t|\text{type} = 1) = \exp\{-\exp(0.3852 \cdot 1)0.0017 t^{0.778}\} = \exp(-0.0025 t^{0.778})$$

and the survival function corresponding to the laparoscopic technique as

$$S(t|\text{type} = 0) = \exp\{-\exp(0.3852 \cdot 0)0.0017 t^{0.778}\} = \exp(-0.0017 t^{0.778})$$

Using Equation 6, the estimated median survival time for patients operated by the open surgical technique is 1378 days, while the estimated median survival time for patients operated by the laparoscopic technique is 2261 days. These results point to a poorer survival experience for patients operated by the open surgical technique. For assessment of Weibull distribution assumption, the log minus log of the Kaplan-Meier survival estimates versus the log of survival time is plotted in Fig. 1. The black curve represents the group of patients operated by the open surgical technique and the gray curve is assigned to the group of patients operated by the laparoscopic surgical technique. Both curves look reasonably straight, what suggests that the Weibull assumption is hold, moreover the lines appear to be parallel (i.e. they have the same slope) suggesting that the proportional hazards is hold.

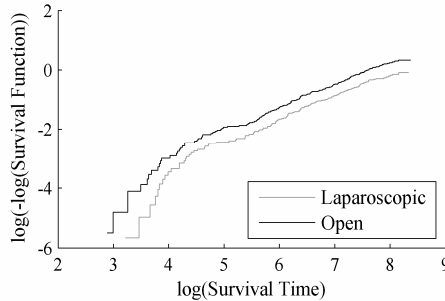


Fig. 1. Graph of the log negative log Kaplan-Meier survival estimates versus log of survival time for assessment of Weibull and proportional hazards assumption for patients with resection of colon.

5 Comparison of Survival Data of Patients with Rectum Surgery

Weibull proportional hazards model, given by Equation 4, containing single covariate was fitted to the survival data of 342 patients who underwent the surgery resection of rectum. Results of this Weibull fit are presented in Table 2.

Table 2. Estimated coefficients, standard error, z-score, two-tailed p-value for the Weibull regression model containing type of surgical technique for patients with resection of colon.

Variable	Value	Std. Err.	z	$p > z $
Type	0.256	0.1772	1.45	0.1479
Constant	7.821	0.1156	67.68	0.0000
ln(sigma)	0.196	0.0624	3.15	0.0017
sigma	1.217			

The p -value for the Wald statistics for the type of surgical technique greater than 0.05 suggests; that the effect of the type of surgical technique is not statistically significant.

Using the same approach as in Sect. 4 we get the values for shape and scale parameter as $\lambda = 1/1.217 = 0.822$ and $\gamma = \exp(-7.821/1.217) = 0.0016$.

The hazard function for the Weibull regression model is then $h(t|\mathbf{x}) = 0.822 \cdot 0.0016 t^{0.822-1} = 0.0013 t^{-0.178}$

The survival function is the same for both of the types of surgical technique $S(t) = \exp\{-0.0016 t^{0.822}\}$

The estimated median survival time for all patients is 1618 days. Fig. 2 shows the log minus log of the Kaplan-Meier survival estimates versus the log of survival time. Because the curve looks reasonably straight, we may conclude that the Weibull assumption is hold.

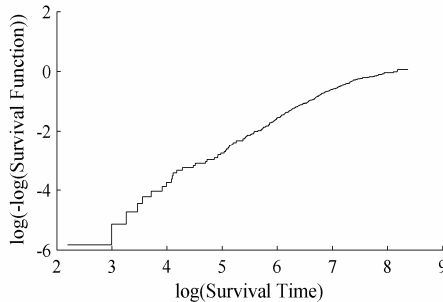


Fig. 2. Graph of the log negative log Kaplan-Meier survival estimates versus log of survival time for assessment of Weibull assumption for patients with resection of rectum.

6 Conclusion

The Weibull proportional hazards regression model has been used for analysis of right-censored medical survival data of patients who underwent colectomy at the University Hospital in Ostrava, to compare survival under two different types of surgical technique for resection of colon or rectum in order to answer the relevant question, which type of surgical technique to choose to guarantee longer overall survival time. The surgical techniques were compared separately for patients with resection of colon and for patients with resection of rectum.

Regarding the surgery of colon the model led to the result that the effect of the type of surgical technique is statistically significant and that the estimated median survival time for patients operated by the open surgical technique is 1378 days, while the estimated median survival time for patients operated by the laparoscopic technique is 2261 days. Therefore we can conclude that the laparoscopic surgical technique guarantees significantly longer survival time in comparison with the open surgical technique for patients who underwent resection of colon.

Concerning the comparison of surgical techniques for surgery of rectum, the model led to the result that the type of surgical technique is not associated with survival time. As a result, we conclude that there is no statistically significant difference between survival times of different surgical techniques for patients who underwent resection of rectum.

Models containing more covariates such as age, gender, bmi, blood loss, stage and grading of cancer are topics of our future research.

This article is a part of paper presented on the European Safety and Reliability Conference in Amsterdam 2013.

References

1. Abraham, N.S. et al.: Meta-analysis of short-term outcomes after laparoscopic resection for colorectal cancer, *Br. J. Surg.* 91, (2004), 1111-1124
2. Bonjer, H.J et al.: Laparoscopically assisted vs open colectomy for colon cancer. *Arch. Surg.* 142, (2007), 298-303
3. Fleshman, J. et al.: Laparoscopic colectomy for cancer is not inferior to open surgery based on 5-year data from the COST study group trial. *Ann. Surg.*, 246, (2007), 655-662
4. Jackson, T.D. et al.: Laparoscopic versus open resection for colorectal cancer: a metaanalysis of oncologic outcomes. *J. Am. Coll. Surg.* 204, (2007), 439-446
5. Kitano S. et al.: A multicenter study on laparoscopic surgery for colorectal cancer in Japan. *Surg. Endos.* 20, (2006), 1348-1352
6. Schwenk, W. et al.: Short term benefits for laparoscopic colorectal resection. *Cochrane database of systematic reviews*, Issue 2.: art. No.CD003145, (2005)
7. Veldkamp R. et al.: Laparoscopic resection of colon cancer: Consensus of the European Association of Endoscopic Suregry. *Surg. Endosc.* 18, (2004), 1163-1185
8. Collett, D.: *Modeling survival data in medical research*. Second edition. London: Chapman & Hall/CRC, (2003)
9. Harrell, E.F.: *Regression Modeling Strategies With Application to Linear Models, Logistic Regression and Survival Analysis*. New York: Springer-Verlag, (2001)
10. Hosmer, D.W. et al.: *Applied Survival Analysis, Regression Modeling of Time-to-Event Data*. Hoboken: Wiley, (2008)
11. Kleinbaum, D. G. & Klein, M.: *Survival Analysis: A Self learning text*. New York: Springer, (1996)

Tangential Fields in Mathematical Modelling of Optical Diffraction Problems

Jiří Krčák and Jaroslav Vlček

Department of Mathematics and Descriptive Geometry, VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33, Ostrava-Poruba
jiri.krcek@vsb.cz

Abstract. Optical diffraction for periodical interface belongs to relatively fewer exploited application of boundary integral equations method. Our contribution presents the formulation of diffraction problem based on vector tangential fields, for which the periodical Green function of Helmholtz equation is of key importance. There are discussed properties of obtained boundary operators with singular kernel and a numerical implementation is proposed.

Keywords: optical diffraction, boundary integral method, tangential fields

1 Introduction

The geometrical and material optimization of the sensors, switching elements and many other devices depends on the accurate control of their parameters. Besides less or more complicated experiments, theoretical studies are carried out including mathematical models of electromagnetic wave interaction with geometrically or material-wise modulated media. Generally, these models consist in the solving of Maxwell equations with appropriate boundary conditions.

In the last two decades, there were published numerous works treating of optical diffraction in periodical structures - see [1] and references therein. One of relatively new approaches is based on Boundary Integral Equations (BIE), theoretical background of which is referred e.g. in [2]. In this article, we aim to show the especial integral formulation of the boundary problem for system of Maxwell equations. To this purpose, we introduce tangential vector fields and study the properties of derived integral operators.

2 Formulation of the problem

Let's denote $\mathbf{X} = (x_1, x_2, x_3) \in \mathbb{R}^3$ and further $S : x_3 = f(x_1)$ a surface which we consider to be smooth with normal vector $\boldsymbol{\nu}$ and periodically modulated in coordinate x_1 with period Λ and uniform in the x_2 direction, see Fig.1.

The interface S divides the space into two semi-infinite homogeneous regions $\Omega^{(1)} = \{\mathbf{X} \in \mathbb{R}^3, x_3 > f(x_1)\}$, $\Omega^{(2)} = \{\mathbf{X} \in \mathbb{R}^3, x_3 < f(x_1)\}$ with constant relative permittivities $\varepsilon^{(1)} \neq \varepsilon^{(2)}$, $\varepsilon^{(1)} \in \mathbb{R}$ and $\varepsilon^{(2)} \in \mathbb{C}$, $\text{Re}(\varepsilon^{(2)}) > 0$, $\text{Im}(\varepsilon^{(2)}) \geq 0$, and, relative permeabilities $\mu^{(1)} = \mu^{(2)} = 1$ (materials are magnetically neutral).

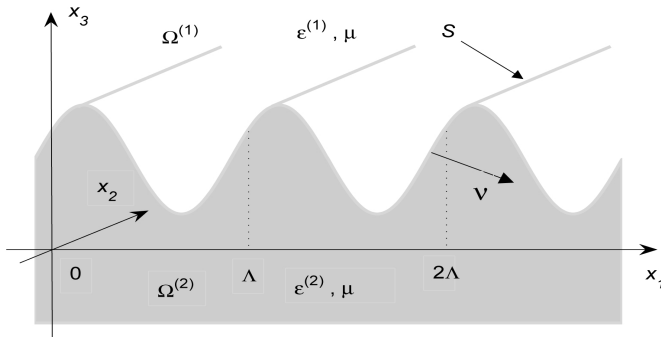


Fig. 1. Structure of regions with common periodical boundary.

We aim to solve optical diffraction problem for monochromatic plane wave with wavelength λ , i.e. with wave number $k_0 = 2\pi/\lambda$ that is incoming from $\Omega^{(1)}$ under the angle of incidence θ measured from x_3 direction. We seek for space-dependent amplitudes $\mathbf{E}^{(j)} = \mathbf{E}|_{\Omega^{(j)}}$, $\mathbf{H}^{(j)} = \mathbf{H}|_{\Omega^{(j)}}$ of the electromagnetic field intensity vectors $\mathbf{E}(\mathbf{X})e^{-i\omega t}$, $\mathbf{H}(\mathbf{X})e^{-i\omega t}$, where $\omega = c/\lambda$ and c represents the light velocity in the free space. Especially, we suppose the TM polarization of the incident wave, for which $\mathbf{E}^{(j)} = (E_1^{(j)}, 0, E_3^{(j)})$, $\mathbf{H}^{(j)} = (0, H_2^{(j)}, 0)$.

Therefore, the Maxwell problem leads to the Helmholtz equations for the scalar components $H_2^{(j)}(\mathbf{X})$,

$$\Delta H_2^{(j)} + k_0^2 \varepsilon^{(j)} H_2^{(j)} = 0 \quad \text{on } \Omega^{(j)}, \quad j = 1, 2. \tag{1}$$

The tangential components of the fields are continuous on the boundary, i.e.

$$\boldsymbol{\nu} \times (\mathbf{E}^{(1)} - \mathbf{E}^{(2)}) = \mathbf{o}, \quad \boldsymbol{\nu} \times (\mathbf{H}^{(1)} - \mathbf{H}^{(2)}) = \mathbf{o} \quad \text{on } S. \tag{2}$$

For the far fields, the well-known Sommerfeld's radiation convergence conditions hold that enable to consider the problem on the common interface S only [3].

The incident field at zero diffraction order is characterized by the relation

$$\mathbf{H}_0^{(1-)} = e^{-i\omega t} e^{i(\alpha x_1 + \beta_0^{(1-)} x_3)} \mathbf{e}_2, \quad \mathbf{e}_2 = (0, 1, 0), \tag{3}$$

where $\alpha = k_0 \sqrt{\varepsilon^{(1)}} \sin \theta$ and $\beta_0^{(1-)}$ is the propagation constant defined below.

This optical beam is diffracted into reflected wave in $\Omega^{(1)}$ and transmitted one in $\Omega^{(2)}$, which are represented by countable sets of modes with wave vectors

$$\mathbf{k}_m^{(j\pm)} = (\alpha_m, 0, \beta_m^{(j\pm)}), \quad \alpha_m = \alpha + 2\pi m / \Lambda, \quad (\beta_m^{(j\pm)})^2 = k_0^2 \varepsilon^{(j)} - \alpha_m^2, \quad m \in \mathbb{Z}. \tag{4}$$

The sign in superscript denotes propagation direction with respect to the x_3 axis orientation: "+" means the forward wave (reflected), "-" the backward one (incident, transmitted). For example $\beta_m^{(j-)} < 0$, if $\beta_m^{(j-)} \in \mathbb{R}$, or, $\text{Im}(\beta_m^{(j-)}) < 0$ otherwise with respect to radiation conditions and chosen convention $e^{-i\omega t}$ – see equation (3). In what follows stay 1 for 1+ and 2 for 2-.

Denoting $\mathbf{x} = (x_1, x_3)$, $\mathbf{y} = (y_1, y_3)$, the periodical fundamental solution of the Helmholtz equation in $\Omega^{(j)}$ can be written as [4]

$$\Psi^{(j)}(\mathbf{x}, \mathbf{y}) = \frac{1}{2i\Lambda} \sum_{m=-\infty}^{\infty} \Psi_m^{(j)}(\mathbf{x}, \mathbf{y}), \quad \Psi_m^{(j)}(\mathbf{x}, \mathbf{y}) = \frac{1}{\beta_m^{(j)}} e^{i(\alpha_m(x_1 - y_1) + \beta_m^{(j)}|x_3 - y_3|)}. \tag{5}$$

In further considerations we exploit following well-known property of the functions $\Psi^{(j)}$.

Theorem 1. *For both of the function $\Psi^{(j)}(\mathbf{x}, \mathbf{y})$ defined by (5) the difference (6) is continuous in \mathbb{R}^2 .*

$$\Psi^{(j)}(\mathbf{x}, \mathbf{y}) - \frac{1}{2\pi} \ln \frac{1}{\|\mathbf{x} - \mathbf{y}\|} \tag{6}$$

3 Boundary integral equations

The aim of this section is to formulate boundary integral equations for tangential fields

$$\mathbf{J} = \boldsymbol{\nu} \times \mathbf{E}^{(1)} = \boldsymbol{\nu} \times \mathbf{E}^{(2)}, \quad \mathbf{I} = -\boldsymbol{\nu} \times \mathbf{H}^{(1)} = -\boldsymbol{\nu} \times \mathbf{H}^{(2)}, \tag{7}$$

where $\boldsymbol{\nu} = (f', 0, -1) / \sigma$ with $\sigma = \sqrt{1 + f'^2}$ is an unit normal vector of the reduced boundary $S : x_3 = f(x_1)$ oriented as shown in Fig.1.

Similarly, $\boldsymbol{\tau} = (1, 0, f') / \sigma$ represents an unit tangential vector of S .

Thus, on the boundary we can write

$$\mathbf{J} = -J_2 \mathbf{e}_2, \quad \text{where } J_2 = \boldsymbol{\tau} \cdot \mathbf{E}^{(1)} = \boldsymbol{\tau} \cdot \mathbf{E}^{(2)}, \tag{8}$$

and,

$$\mathbf{I} = \sigma I_1 \boldsymbol{\tau} = I_\tau \boldsymbol{\tau}, \quad \text{where } I_\tau = \sigma I_1 = -H_2^{(1)} = -H_2^{(2)}. \tag{9}$$

For boundary points $\boldsymbol{\xi} = (\xi_1, \xi_3)$, $\boldsymbol{\eta} = (\eta_1, \eta_3)$ on the interface $S : \eta_3 = f(\eta_1)$, $\eta_1 \in \langle 0, \Lambda \rangle$ we obtain following system of boundary integral equations [5]

$$\begin{aligned}
 J_2(\boldsymbol{\xi}) &= -J_0(\boldsymbol{\xi}) - ik_0 \boldsymbol{\tau}_\xi \cdot \int_S I_\tau \boldsymbol{\tau}_\eta (\Psi^{(1)} - \Psi^{(2)}) dl_\eta \\
 &- \frac{1}{ik_0} \boldsymbol{\tau}_\xi \cdot \int_S \frac{1}{\sigma} \frac{dI_\tau}{d\eta_1} \nabla_\eta \left(\frac{1}{\varepsilon^{(1)}} \Psi^{(1)} - \frac{1}{\varepsilon^{(2)}} \Psi^{(2)} \right) dl_\eta + \boldsymbol{\nu}_\xi \cdot \int_S J_2 \nabla_\eta (\Psi^{(1)} - \Psi^{(2)}) dl_\eta,
 \end{aligned} \tag{10}$$

$$I_\tau(\boldsymbol{\xi}) = -I_0(\boldsymbol{\xi}) - ik_0 \int_S J_2(\varepsilon^{(1)} \Psi^{(1)} - \varepsilon^{(2)} \Psi^{(2)}) dl_\eta + \int_S I_\tau \boldsymbol{\nu}_\eta \cdot \nabla_\eta (\Psi^{(1)} - \Psi^{(2)}) dl_\eta, \tag{11}$$

where

$$J_0(\boldsymbol{\xi}) = -\mathbf{e}_2 \cdot (\boldsymbol{\nu}_\xi \times \mathbf{E}_0^{(1-)}) = \boldsymbol{\tau}_\xi \cdot \mathbf{E}_0^{(1-)}, \quad I_0(\boldsymbol{\xi}) = \boldsymbol{\tau}_\xi \cdot (\boldsymbol{\nu}_\xi \times \mathbf{H}_0^{(1-)}) = -H_{0,2}^{(1-)}, \tag{12}$$

thereby $\mathbf{E}_0^{(1-)}, \mathbf{H}_0^{(1-)}$ represent the incident wave in $\Omega^{(1)}$.

When deriving these equations it is necessary to study properties of integral operators

$$\int_S g(\boldsymbol{\eta}) \psi(\mathbf{x}, \boldsymbol{\eta}) dl_\eta, \quad \int_S g(\boldsymbol{\eta}) \frac{\partial \psi(\mathbf{x}, \boldsymbol{\eta})}{\partial \boldsymbol{\nu}} dl_\eta, \quad \int_S g(\boldsymbol{\eta}) \nabla_\eta \psi(\mathbf{x}, \boldsymbol{\eta}) dl_\eta \tag{13}$$

with the kernel

$$\psi(\mathbf{x}, \boldsymbol{\eta}) = \frac{1}{2\pi} \ln \frac{1}{\|\mathbf{x} - \boldsymbol{\eta}\|} \tag{14}$$

when crossing from the inner point \mathbf{x} to the boundary point $\boldsymbol{\xi}$ in the normal direction (the superscript (j) is omitted for simplicity).

Whereas the first and the second of these are the well-known single and double layer potentials, the third is worth to mention.

Theorem 2. *Let $\psi(\mathbf{x}, \boldsymbol{\eta})$ is the function (14) and S is smooth boundary of the domain $\Omega \subset \mathbb{R}^2$ with unit outward normal $\boldsymbol{\nu}$. If $g \in C(S)$, then*

$$\lim_{\mathbf{x} \rightarrow \boldsymbol{\xi}} \int_S g(\boldsymbol{\eta}) \nabla_\eta \psi(\mathbf{x}, \boldsymbol{\eta}) dl_\eta = \int_S g(\boldsymbol{\eta}) \nabla_\eta \psi(\boldsymbol{\xi}, \boldsymbol{\eta}) dl_\eta \pm \frac{1}{2} g(\boldsymbol{\xi}) \boldsymbol{\nu}(\boldsymbol{\xi}), \tag{15}$$

where $\boldsymbol{\xi} \in S$, minus holds for $\mathbf{x} \in \Omega$ and plus for $\mathbf{x} \in \mathbb{R}^2 \setminus \bar{\Omega}$.

4 Operator form

Let $\boldsymbol{\pi} : \langle 0, 2\pi \rangle \rightarrow \mathbb{R}^2$, $\boldsymbol{\pi}(t) = (p(t), q(t))$ be a parametrization of the boundary S . For the boundary points we have $\boldsymbol{\xi} = \boldsymbol{\pi}(s)$, $\boldsymbol{\eta} = \boldsymbol{\pi}(t)$, $s, t \in \langle 0, 2\pi \rangle$ with corresponding unit normal vector $\boldsymbol{\nu}(t) = (\nu_1(t), \nu_3(t)) = (q'(t), -p'(t))/\nu(t)$ and unit tangential vector $\boldsymbol{\tau}(t) = (p'(t), q'(t))/\nu(t)$, where $\nu(t) = \sqrt{p'(t)^2 + q'(t)^2}$.

In the integral operators kernels the fundamental solution (5) of the Helmholtz equation takes place, hence the system (10), (11) can be written in operator form

$$\begin{bmatrix} \mathcal{V}_1 + \mathcal{V}_2 \mathcal{I} - \mathcal{V}_3 \\ \mathcal{I} - \mathcal{V}_4 & \mathcal{V}_5 \end{bmatrix} \begin{bmatrix} I_\tau \\ J_2 \end{bmatrix} = \begin{bmatrix} -J_{2,0} \\ -I_{\tau,0} \end{bmatrix}, \tag{16}$$

where \mathcal{I} is the identity operator. The operators $\mathcal{V}_1, \dots, \mathcal{V}_5$ containing differences of the fundamental solutions or gradients of these are derived in [7] in detail.

The right-hand terms of (16) are simply obtained by parametrization of incident fields (12).

5 Properties of boundary integral operators

The structure of operators in (16) make us to discuss properties of integral operators kernels, which are written as differences $c_1\Psi^{(1)}(s, t) - c_2\Psi^{(2)}(s, t)$, or their gradients, where c_1, c_2 are generally complex constants. Because for $s \neq t$ this expression represents a continuous function, it suffices to analyse the singular case for $s = t$.

Theorem 3. *Let $c_1, c_2 \in \mathbb{C}$. Then for $s = t$ the functions*

$$c_1\Psi^{(1)}(s, t) - c_2\Psi^{(2)}(s, t), \quad \nabla_t \left(c_1\Psi^{(1)}(s, t) - c_2\Psi^{(2)}(s, t) \right) \tag{17}$$

are continuous for $c_1 = c_2$ and these have singularity of logarithmic type for $c_1 \neq c_2$.

The particular manner how to evaluate singular integrals depends on the choice of numerical method. The following theorems show one of possible methods - see [6], where also the proofs can be found ($\mathbb{Z}^* = \mathbb{Z} - \{0\}$).

Theorem 4. *Let $\pi : \langle 0, 2\pi \rangle \rightarrow \mathbb{R}^2$ is a parametrization that satisfies*

$$p(0) = 0, \quad p(2\pi) = A, \quad q(0) = q(2\pi), \quad p(t + 2\pi) = p(t) + A, \quad q(t + 2\pi) = q(t).$$

Then

$$\ln \|\pi(s) - \pi(t)\| = \ln \left| 2 \sin \frac{s-t}{2} \right| = - \sum_{m \in \mathbb{Z}^*} \frac{e^{-im(s-t)}}{2|m|}. \tag{18}$$

Theorem 5. *The series (19) is absolutely convergent for arbitrary $s, t \in \langle 0, 2\pi \rangle$.*

$$\sum_{m \in \mathbb{Z}^*} \left\{ \Psi_m^{(j)}(s, t) - \frac{1}{2\pi} \frac{e^{-im(s-t)}}{2|m|} \right\} \tag{19}$$

These properties together with Theorem 1 allow us to split the fundamental solution as

$$\Psi^{(j)}(s, t) = \Psi_r^{(j)}(s, t) + \psi(s, t), \tag{20}$$

where

$$\Psi_r^{(j)}(s, t) = \Psi_0^{(j)}(s, t) + \sum_{m \in \mathbb{Z}^*} \left\{ \Psi_m^{(j)}(s, t) - \frac{1}{2\pi} \frac{e^{-im(s-t)}}{2|m|} \right\}, \quad (21)$$

$$\psi(s, t) = \frac{1}{2\pi} \ln \left| 2 \sin \frac{s-t}{2} \right|. \quad (22)$$

In numerical implementations we work separately with regular integral kernels and with singular integrals which can be evaluated analytically.

6 Conclusion

The presented formulation of diffraction problem represents appropriate background of numerical solution by the Boundary Elements Method (BEM). Obtained values of tangential fields enable to compute electromagnetic intensities on the boundary and above all to extrapolate these out of the boundary.

The specific problem to discuss is the choice of basis functions; trigonometric polynomials have been used in [3], for instance. For our further work we prefer piecewise linear boundary elements. The other problem of numerical implementation is the choice of appropriate integration method, the trapezoidal rule seems to be sufficient for required precision.

References

1. Bao, G., Cowsar, L. and Masters, W.: *Mathematical modeling in optical science*. SIAM, Philadelphia, 2001.
2. Nedelec, J.C., Starling, F.: Integral equation methods in a quasi-periodic diffraction problem for the time-harmonic Maxwell's equations. *SIAM, J. Math. Anal.* **22** (1991), 1679-1701.
3. Kleemann, B.H., Mitreiter, A. and Wyrowski, F.: Integral equation method with parametrization of grating profile. *Theory and Experiments. J. Mod. Opt.* **43** (1996), No. 7, 1323-1349.
4. Linton, C.M.: The Green's function for the two-dimensional Helmholtz equation in periodic domains. *J. Eng. Math.* **33** (1998), 377-402.
5. Dobson, D.C. and Cox, J.A.: An integral equation method for bi-periodic diffraction structures. In: *Proc. of SPIE*, Vol. 1545 (1991), 106-113.
6. Žídek, A., Vlček, J. and Krček, J.: Solution of diffraction problems by boundary integral equations. In: *Proc. of 11th International Conference APLIMAT 2012, Febr. 7-9, 2012, Bratislava, Slovak Republic*, publ. by Faculty of Mechanical Engineering, Slovak University of Technology, Bratislava 2012, 221-229.
7. Krček, J., Vlček, J. and Žídek, A.: Tangential fields in optical diffraction problems. In: *Proc. of Seminar PANM16, June 3-8, 2012, Dolní Maxov*, publ. by Institute of Mathematics Academy of Sciences of the Czech Republic, Prague 2013, 124-129.

Parallel implementation of fast boundary element method

Michal Merta

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
`michal.merta@vsb.cz`

Abstract. In this paper we present a new library for parallelization of boundary element method based solvers. We use the fast multipole method for the sparsification of system matrices. To distribute the workload among processes as evenly as possible, we use the cyclic graph decomposition. The implementation is based on C++ code vectorized by Vc library and parallelized using MPI.

1 Introduction

Using boundary element method for the solution of engineering problems we reduce a dimension of a problem from d to $d - 1$. A problem formulation is reduced to a boundary of a computational domain, which is particularly useful in the case of unbounded domains. This approach not only significantly reduces a number of unknowns (when compared to the finite element method), but also the time necessary for a mesh generation. However, the method produces dense matrices requiring $\mathcal{O}(N^2)$ operations for assembling and the same amount of operation per matrix-vector multiplication in iterative solvers. The high computational complexity and memory requirements limit the usage of the method for real world problems, therefore some kind of parallelization combined with a method for matrix sparsification has to be employed. In our work we use the fast multipole method (FMM) for the sparsification of system matrix blocks. The distribution of matrix blocks among processors is based on the cyclic graph decomposition to ensure as evenly distributed workload as possible.

The outline of this paper is as follows: in Section 2 we describe the model problem and the associated boundary element formulation, in Section 3 we present the basic idea of the fast multipole method. The next section deals with an efficient implementation and parallelization. Finally, in Section 5 we present the results of numerical experiments.

2 Model problem

For the sake of simplicity let us consider the following Dirichlet boundary value problem for the Laplace operator:

$$\begin{cases} -\Delta u = 0 & \text{in } \Omega, \\ \gamma^0 u = g & \text{on } \partial\Omega, \end{cases}$$

with $u \in H^1(\Omega)$ being an unknown function and $g \in H^{1/2}(\partial\Omega)$ being a prescribed Dirichlet datum. The solution of this problem can be obtained by BEM, i.e. using the representation formula

$$u(x) = (\tilde{V}\gamma^1 u)(x) - (W\gamma^0 u)(x), \quad x \in \Omega, \tag{1}$$

where

$$(\tilde{V}s)(x) := \int_{\partial\Omega} G(x, y)s(y) \, ds_y, \quad (Wu)(x) := \int_{\partial\Omega} \frac{\partial G(x, y)}{\partial n_y} u(y) \, ds_y,$$

are the single and double layer potentials, respectively. By $G(x, y) := \frac{1}{4\pi} \frac{1}{\|x-y\|}$ we denote the fundamental solution of the Laplace equation in 3D.

Applying the internal trace operator $\gamma^0 : H^1(\Omega) \rightarrow H^{1/2}(\partial\Omega)$ to both sides of Equation (1) we obtain the boundary integral equation

$$u(x) = (V\gamma^1 u)(x) - \left(-\frac{1}{2}u(x) + K(u)(x) \right), \quad x \in \partial\Omega, \tag{2}$$

where $\gamma^1 u := \frac{\partial u}{\partial n}$. Operators $V := \gamma^0 \circ \tilde{V} : H^{-1/2}(\partial\Omega) \rightarrow H^{1/2}(\partial\Omega)$ and $K : H^{1/2}(\partial\Omega) \rightarrow H^{1/2}(\partial\Omega)$ are given by

$$(Vs)(x) := \int_{\partial\Omega} G(x, y)s(y) \, ds_y, \quad (Ku)(x) := \int_{\partial\Omega} \frac{\partial G(x, y)}{\partial n_y} u(y) \, ds_y.$$

For the discretization we use the Galerkin method with piece-wise constant basis and testing functions. After a triangulation $T := \cup_{\ell=1}^N \tau_\ell$ we obtain the following system of linear equations

$$V_h t = \left(\frac{1}{2}M_h + K_h \right) g.$$

The matrices V_h, K_h of the single layer and double layer potential, respectively, are given by:

$$V_h^{ij} := \langle V\psi_j, \psi_i \rangle_{\partial\Omega}, \quad K_h^{ij} := \langle K\psi_j, \psi_i \rangle_{\partial\Omega},$$

and M_h is the diagonal matrix with entries $m_{ii} = |\tau_i|$. Because of a nonlocality of the kernel function G , the matrices V_h and K_h are fully populated.

3 Fast multipole method

The main idea of the fast multipole method [1, 5, 6] for BEM matrices sparsification is to leverage the expansion of the kernel by the spherical harmonic functions

$$\frac{1}{\|x - y\|} \approx \sum_{n=0}^p \sum_{m=-n}^n \overline{S_n^m}(y) R_n^m(x), \tag{3}$$

$$R_n^{\pm m}(x) = \frac{1}{(n + m)!} \frac{d^m}{du^m} P_n(u)|_{u=\hat{x}_3} (\hat{x}_1 \pm i\hat{x}_2)^m |x|^n,$$

$$S_n^{\pm m}(y) = (n - m)! \frac{d^m}{du^m} P_n(u)|_{u=\hat{y}_3} (\hat{y}_1 \pm i\hat{y}_2)^m \frac{1}{|y|^{n+1}},$$

$\hat{y}_i = y_i/\|y\|$, $\|x\| < \|y\|$. This leads to

$$\int_{\tau_j} \int_{\tau_i} \frac{1}{\|x - y\|} ds_x ds_y \approx \sum_{n=0}^p \sum_{m=-n}^n \int_{\tau_j} R_n^{\pm m}(x) ds_x \int_{\tau_i} \overline{S_n^m}(y) ds_y,$$

which significantly reduces the computational complexity because the integrals by x and y are now decoupled.

Since the expansion (3) is only valid for $\|x\| < \|y\|$ we use the recursive geometrical bisection to split a computational domain into clusters and to construct a binary cluster tree. The pair of clusters (C_x, C_y) is said to be *admissible* if it satisfies the condition

$$\min \{ \text{diam}C_x, \text{diam}C_y \} \leq \eta \text{dist}(C_x, C_y),$$

otherwise it is called *nonadmissible*. If the pair of clusters is admissible we say that the cluster C_y is in the *farfield* of the cluster C_x and vice versa. Otherwise, they are in their mutual *nearfield*. The admissible cluster pairs correspond to the blocks of matrix approximated by means of FMM.

Using the expansion (3) a matrix-vector multiplication $w = At$ can be evaluated effectively by splitting it into a nearfield and farfield part

$$w_i = \sum_{j \in \text{NF}(i)} A_{ij} t_j + \sum_{n=0}^p \sum_{m=-n}^n \hat{M}_n^m(O, \psi_i) \tilde{L}_n^m(O, \text{FF}(i)),$$

where $\text{NF}(i)$, $\text{FF}(i)$ are the sets of clusters in the nearfield or farfield, respectively, of the cluster containing the element τ_i . By $\hat{M}_n^m(O, \psi_i)$, $\tilde{L}_n^m(O, \text{FF}(i))$ we denote multipole moments and coefficients of a local expansion associated with a given cluster, respectively. Its efficient computation leverages the existing tree structure:

1. *Upward pass* - multipole moments are computed on the finest level of the tree and translated to the higher levels by multipole to multipole (M2M) translations.

2. *Downward pass* - coefficients of a local expansion are computed on the highest possible level by translation of multipole moments (M2L), and translated to the lower levels by local to local translations (L2L).

Since the multipole coefficients depend on the vector t , these tree traversals have to be repeated in each iteration of an iterative solver. For details see [4].

4 Implementation details

The presented method is a part of our newly developed library of parallel BEM based solvers BEM4I. This library is based on C++ and uses MPI for a global communication. A hybrid MPI/OpenMP model is planned for a future development.

To gain a maximum performance of current processors, one has to pay a great attention to the vectorization of the codes. Both the width of vector registers and SIMD instruction sets are being extended and a performance of a program heavily depends on its proper vectorization. However, the compiler is not always capable of automatic loop vectorization. Since the usage of assembly language or intrinsic functions can lead to a confusing and hardly manageable code, we have decided to use a high level C++ vectorization library Vc. It provides API for explicit vectorization, independent of a compiler version or an instruction set of a processor [7]. Using Vc we have vectorized the most computationally demanding parts of system matrix assembly.

To parallelize our code we decompose an underlying mesh into N submeshes and the resulting matrix into corresponding $N \times N$ blocks, each of which is approximated by a sequential fast multipole method. Each of N processes is assigned one diagonal block (these are typically most time and memory consuming within the fast BEM) and $N - 1$ geometrically closely related off-diagonal blocks, thus the total memory consumption for storing the mesh and related structures is minimal (see Fig. 1). With this decomposition we can achieve an optimal parallel computational scalability $\mathcal{O}((n \log n)/N)$ and reasonable memory scalability

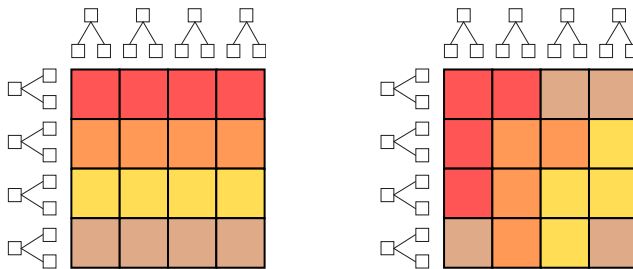


Fig. 1. Comparison of naive matrix distribution (left) and optimal distribution

Table 1. Scalability of the problem solution

Problem dimension	196,608				
Number of cores	7	13	31	73	133
V-matrix assembly [s]	208	133	65	31	19

$\mathcal{O}((n \log n)/\sqrt{N})$. The problem of finding an optimal decomposition corresponds to the problem of decomposition of complete undirected graphs, which is known for N such that it holds

$$\frac{N(N - 1)}{2N} = \frac{p(p - 1)}{2}.$$

Here $p + 1$ is a power of a prime number. A detailed description is given in [2].

5 Numerical experiments

The scalability experiments were carried out on a cluster consisting of nodes equipped with two AMD 2.3 GHz 16-core Interlagos processors. Its floating point unit consists of two 128-bit pipelines, therefore it is capable of executing either single 256-bit AVX instruction, or two 128-bit SSE instructions per cycle.

Fig. 2 depicts times of assembly of the matrix K with a dimension $196,608 \times 196,608$. We can see that the vectorized code is almost four times faster than the non-vectorized one. Table 1 shows the scalability of the non-vectorized matrix V assembly. The difference between vectorized and non-vectorized version is not significant in this case.

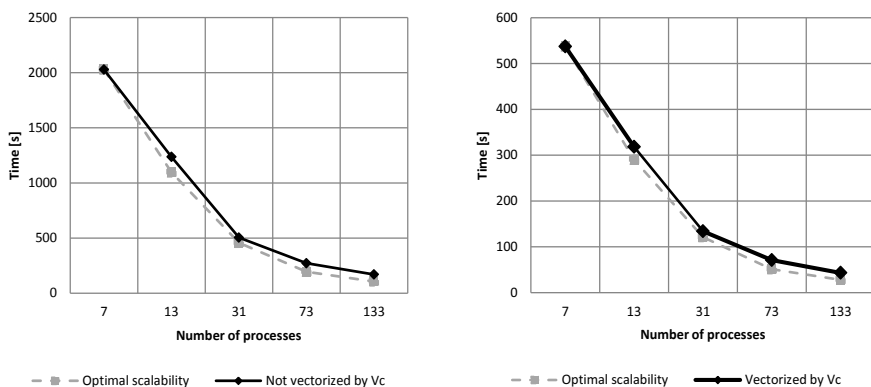


Fig. 2. Scalability of the matrix K assembly for non-vectorized and vectorized code

6 Conclusion

In this paper we presented the new library for parallel fast BEM solvers. The parallelization of the library is based on cyclic decompositions of undirected graphs. Some computational kernels are vectorized using Vc library. In the future work we would like to vectorize routines related to the matrix-vector multiplication in the FMM and implement hybrid MPI/OpenMP parallelization.

Acknowledgements

This work was supported by the European Regional Development Fund in the IT4I Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and by the project SPOMECH - Creating a multidisciplinary R&D team for reliable solution of mechanical problems, reg. no. CZ.1.07/2.3.00/20.0070 within Operational Programme 'Education for competitiveness' funded by Structural Funds of the European Union and state budget of the Czech Republic.

This work received a grant of 100,000 core-hours in the IT4Innovations 1st Internal Access Call.

References

1. Greengard, L., Rokhlin, V.: *A fast algorithm for particle simulations*. J. Comput. Phys., 73, pp. 325–384, 1987.
2. Lukáš, D., Kovář, P., Kovářová, T., Merta, M.: *A parallel fast boundary element method using cyclic graph decompositions*. Submitted to Numerical Algorithms.
3. Merta, M., Lukáš, D.: Parallel implementation of fast boundary element method. In: *Seminar on Numerical Analysis*, Institute of Geonics AS CR, Ostrava, 2013, pp. 90–92.
4. Of, G.: Fast Multipole Method and Applications. In: Schanz, M. Steinbach, O.: *Boundary element analysis: mathematical aspects and applications*, Springer, London, c2007, pp. 135–160.
5. Rokhlin, V.: *Rapid solution of integral equations of classical potential theory*. J. Comput. Phys., 60, pp. 187–207, 1985.
6. Steinbach, O., Rjasanow, S.: *The Fast Solution of Boundary Integral Equations*. Springer, 2007.
7. Vc Web page, <http://code.compeng.uni-frankfurt.de/projects/vc>

Semi-Smooth Newton Method for Contact Problems with Tresca and Coulomb Friction

Kristina Motyčková and Radek Kučera

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{kristina.motyckova, radek.kucera}@vsb.cz

Abstract The contribution deals with contact problems for two elastic bodies with friction. After the description of the problem we present its discretization based on linear or bilinear finite elements. The semi-smooth Newton method is used to find the solution. We present active sets algorithms. The non-symmetric and symmetric case is distinguished with using BiCGSTAB and CGM, respectively. Finally we will arrive at the globally convergent dual implementation of the algorithm.

Keywords: contact problems, semi-smooth Newton method, dual formulation

1 Introduction

Let us consider two homogeneous isotropic elastic bodies represented by bounded domains $\Omega^k \subset \mathbb{R}^2$ with sufficiently smooth boundaries $\partial\Omega^k$, $k = 1, 2$. Each boundary consists of three disjoint parts Γ_u^k , Γ_p^k , and Γ_c^k open in $\partial\Omega^k$ so that $\partial\Omega^k = \bar{\Gamma}_u^k \cup \bar{\Gamma}_p^k \cup \bar{\Gamma}_c^k$ and $\bar{\Gamma}_u^k \neq \emptyset$; The zero displacements are prescribed on Γ_u^k while surface tractions $\mathbf{p}^k \in (L^2(\Gamma_p^k))^2$ act on Γ_p^k . The bodies may get into contact on the *contact interface* given by Γ_c^1 and Γ_c^2 , where we consider three contact conditions: the non-penetration of bodies, the transmission of contact stresses, and the Coulomb friction law. Elastic properties of Ω^k are described by the Lamè constants $\lambda^k, \mu^k > 0$. Finally, assume the volume forces $\mathbf{f}^k \in (L^2(\Omega^k))^2$.

Our aim is to find an equilibrium state of Ω^1 and Ω^2 . By the solution of this problem we mean displacement vector fields $\mathbf{u}^k = (u_1^k, u_2^k)^\top$, $k = 1, 2$, satisfying the equilibrium equations and the Dirichlet and Neumann conditions.

The contact problem with the Coulomb friction may be solved directly, or by using the method of successive approximations (fixed-point approach) where in each step the problem with the Tresca friction is solved (according to a predefined mapping).

The formulation of the contact problem with Coulomb friction results in an *implicit* variational inequality of the elliptic type. If we replace the slip bounds by an à-priori given positive function g defined on Γ_c^1 , we arrive at the contact problem with Tresca friction. This well known problem is represented by the variational inequality of the second type for which there is a unique solution.

2 Discrete problems

After the discretization of our problem we arrive at

$$\mathbf{K}\mathbf{u} - \mathbf{f} + \mathbf{N}^\top \boldsymbol{\lambda}_\nu + \mathbf{T}^\top \boldsymbol{\lambda}_t = \mathbf{0}, \tag{2.1}$$

$$\mathbf{N}\mathbf{u} - \mathbf{d} \leq \mathbf{0}, \boldsymbol{\lambda}_\nu \geq \mathbf{0}, \boldsymbol{\lambda}_\nu^\top (\mathbf{N}\mathbf{u} - \mathbf{d}) = \mathbf{0}, \tag{2.2}$$

$$\left. \begin{aligned} |\lambda_{t,i}| &\leq \mathcal{F}_i \lambda_{\nu,i} \\ |\lambda_{t,i}| &< \mathcal{F}_i \lambda_{\nu,i} \Rightarrow u_{t,i} = 0 \\ |\lambda_{t,i}| &= \mathcal{F}_i \lambda_{\nu,i} \Rightarrow \exists c_t \geq 0 : u_{t,i} = c_t \lambda_{t,i} \end{aligned} \right\} \quad i = 1, \dots, m. \tag{2.3}$$

We use two Lagrange multipliers $\boldsymbol{\lambda}_\nu, \boldsymbol{\lambda}_t \in \mathbb{R}^m$ that are the opposite of the discrete relative normal and tangential contact stresses, respectively. The stiffness matrix and the load vector are represented by \mathbf{K} and \mathbf{f} , respectively. Matrices \mathbf{N} and \mathbf{T} are associated with the contact nodes in normal and tangential direction, respectively. \mathcal{F} stands for the coefficient of the Coulomb friction. Our unknown vector of the nodal displacement is \mathbf{u} . As this is just short report we have omitted the corresponding dimensions.

In order to obtain the discrete contact problem with Tresca friction, we replace $\mathcal{F}_i \lambda_{\nu,i}$ in (2.3) by the entries g_i of $\mathbf{g} \in \mathbb{R}^m, \mathbf{g} \geq \mathbf{0}$.

Next we use the equivalent formulation of the previous problems as the systems of non-smooth equations. We introduce the projection mappings

$$\mathbf{P}_{\mathbb{R}_+^m} : \mathbb{R}^m \mapsto \mathbb{R}_+^m \quad \mathbf{P}_{\Lambda(\mathbf{r})} : \mathbb{R}^m \mapsto \Lambda(\mathbf{r})$$

with $\mathbb{R}_+^m := \{\boldsymbol{\mu} \in \mathbb{R}^m : \boldsymbol{\mu} \geq \mathbf{0}\}, \Lambda(\mathbf{r}) := \{\boldsymbol{\mu} \in \mathbb{R}^m : |\boldsymbol{\mu}| \leq \mathbf{r}\}$, respectively and $\mathbf{r} \in \mathbb{R}^m, \mathbf{r} \geq \mathbf{0}$. The definitions of the components of $\mathbf{P}_{\mathbb{R}_+^m}, \mathbf{P}_{\Lambda(\mathbf{r})}$ are based on the max-function in \mathbb{R}^1 :

$$(\mathbf{P}_{\mathbb{R}_+^m})_i(\boldsymbol{\mu}) = \max\{0, \mu_i\}, \tag{2.4}$$

$$(\mathbf{P}_{\Lambda(\mathbf{r})})_i(\boldsymbol{\mu}) = \max\{0, \mu_i + r_i\} - \max\{0, \mu_i - r_i\} - r_i. \tag{2.5}$$

Denote $\mathbf{y} := (\mathbf{u}^\top, \boldsymbol{\lambda}_\nu^\top, \boldsymbol{\lambda}_t^\top)^\top \in \mathbb{R}^{2n+2m}$ and consider a parameter $\rho > 0$. The discrete contact problem with Coulomb friction is equivalent to the equation

$$\mathbf{G}(\mathbf{y}) = \mathbf{0} \tag{2.6}$$

where $\mathbf{G} : \mathbb{R}^{2n+2m} \mapsto \mathbb{R}^{2n+2m}$ is defined by

$$\mathbf{G}(\mathbf{y}) := \begin{pmatrix} \mathbf{K}\mathbf{u} - \mathbf{f} + \mathbf{N}^\top \boldsymbol{\lambda}_\nu + \mathbf{T}^\top \boldsymbol{\lambda}_t \\ \boldsymbol{\lambda}_\nu - \mathbf{P}_{\mathbb{R}_+^m}(\boldsymbol{\lambda}_\nu + \rho(\mathbf{N}\mathbf{u} - \mathbf{d})) \\ \boldsymbol{\lambda}_t - \mathbf{P}_{\Lambda\mathbb{P}_{\mathbb{R}_+^m}}(\boldsymbol{\lambda}_\nu + \rho(\mathbf{N}\mathbf{u} - \mathbf{d}))(\boldsymbol{\lambda}_t + \rho\mathbf{T}\mathbf{u}) \end{pmatrix}.$$

2.1 Algorithms

As we use the semi-smooth Newton method (SSNM), our algorithms are based on the following iterative scheme:

$$\mathbf{F}^o(\mathbf{y}^{(k-1)})\mathbf{y}^{(k)} = \mathbf{F}^o(\mathbf{y}^{(k-1)})\mathbf{y}^{(k-1)} - \mathbf{F}(\mathbf{y}^{(k-1)}), \quad k = 1, 2, \dots, \tag{2.7}$$

where $\mathbf{F} : \mathbb{R}^{2n+2m} \mapsto \mathbb{R}^{2n+2m}$ is slantly differentiable and $\mathbf{F}^o(\mathbf{y})$ is a slanting function to \mathbf{F} at $\mathbf{y} \in \mathbb{R}^{2n+2m}$. It is well known that sequence $\{\mathbf{y}^{(k)}\}$ generated by (2.7) converges superlinearly to the solution of $\mathbf{F}(\mathbf{y}) = \mathbf{0}$ when the initial iterate $\mathbf{y}^{(0)} \in \mathbb{R}^{2n+2m}$ is a sufficiently accurate approximation of the solution.

We will show an implementation of SSNM that is equivalent to an active set algorithm. Firstly, let us introduce notations. Let $\mathcal{M} = \{1, 2, \dots, m\}$ be the set of all indices and let $\mathbf{y} = (\mathbf{u}^\top, \boldsymbol{\lambda}_\nu^\top, \boldsymbol{\lambda}_t^\top)^\top \in \mathbb{R}^{2n+2m}$ be given. The active set $\mathcal{A}_\nu := \mathcal{A}_\nu(\mathbf{y})$ corresponding to the non-penetration condition is defined by

$$\mathcal{A}_\nu(\mathbf{y}) = \{i \in \mathcal{M} : \lambda_{\nu,i} + \rho(\mathbf{N}\mathbf{u} - \mathbf{d})_i > 0\}$$

and the respective inactive set is its complement $\mathcal{I}_\nu := \mathcal{I}_\nu(\mathbf{y}) = \mathcal{M} \setminus \mathcal{A}_\nu(\mathbf{y})$. For $S \subseteq \mathcal{M}$ we introduce the diagonal matrix

$$\mathbf{D}_S = \text{diag}(s_1, \dots, s_m) \in \mathbb{R}^{m \times m}, \quad s_i = \begin{cases} 1 & \text{for } i \in S, \\ 0 & \text{for } i \notin S. \end{cases}$$

Non-symmetric case. First we propose the algorithm for direct solving discrete contact problems with Coulomb friction. To this end we introduce two inactive sets $\mathcal{I}_t^+ := \mathcal{I}_t^+(\mathbf{y})$, $\mathcal{I}_t^- := \mathcal{I}_t^-(\mathbf{y})$ corresponding to the condition of Coulomb friction:

$$\begin{aligned} \mathcal{I}_t^+(\mathbf{y}) &= \{i \in \mathcal{M} : \lambda_{t,i} + \rho(\mathbf{T}\mathbf{u})_i > \mathcal{F}_i(\lambda_{\nu,i} + \rho(\mathbf{N}\mathbf{u} - \mathbf{d})_i)^+\}, \\ \mathcal{I}_t^-(\mathbf{y}) &= \{i \in \mathcal{M} : \lambda_{t,i} + \rho(\mathbf{T}\mathbf{u})_i < -\mathcal{F}_i(\lambda_{\nu,i} + \rho(\mathbf{N}\mathbf{u} - \mathbf{d})_i)^+\} \end{aligned}$$

where $(\lambda_{\nu,i} + \rho(\mathbf{N}\mathbf{u} - \mathbf{d}))_i^+ = (\mathbf{P}_{\mathbb{R}_+^m})_i(\lambda_\nu + \rho(\mathbf{N}\mathbf{u} - \mathbf{d}))$. The respective active set is $\mathcal{A}_t := \mathcal{A}_t(\mathbf{y}) = \mathcal{M} \setminus (\mathcal{I}_t^+(\mathbf{y}) \cup \mathcal{I}_t^-(\mathbf{y}))$. After suitable adjustments we arrive at the following algorithm.

Algorithm *ActiveSetCoulomb1*

- (0) Set $k := 1$, $\rho > 0$, $\varepsilon_u > 0$, $\mathbf{u}^{(0)} \in \mathbb{R}^{2n}$, $\boldsymbol{\lambda}_\nu^{(0)}, \boldsymbol{\lambda}_t^{(0)} \in \mathbb{R}^m$.
- (1) Define the active and inactive sets at $\mathbf{y} = ((\mathbf{u}^{(k-1)})^\top, (\boldsymbol{\lambda}_\nu^{(k-1)})^\top, (\boldsymbol{\lambda}_t^{(k-1)})^\top)^\top$:

$$\mathcal{A}_\nu = \mathcal{A}_\nu(\mathbf{y}), \mathcal{I}_\nu = \mathcal{I}_\nu(\mathbf{y}), \mathcal{A}_t = \mathcal{A}_t(\mathbf{y}), \mathcal{I}_t^+ = \mathcal{I}_t^+(\mathbf{y}), \mathcal{I}_t^- = \mathcal{I}_t^-(\mathbf{y}).$$

- (2) Solve:

$$\begin{aligned} \left(\begin{array}{c|c|c} \mathbf{K} & \mathbf{N}^\top & \mathbf{T}^\top \\ \mathbf{D}_{\mathcal{A}_\nu} \mathbf{N} & \mathbf{D}_{\mathcal{I}_\nu} & \mathbf{0} \\ \rho \mathbf{D}_{\mathcal{A}_t} \mathbf{T} & \mathcal{F}(\mathbf{D}_{\mathcal{I}_t^-} - \mathbf{D}_{\mathcal{I}_t^+}) (\rho \mathbf{D}_{\mathcal{I}_\nu} - \mathbf{D}_{\mathcal{A}_\nu}) & -\mathbf{D}_{\mathcal{I}_t^+ \cup \mathcal{I}_t^-} \end{array} \right) \begin{pmatrix} \mathbf{u}^{(k)} \\ \boldsymbol{\lambda}_\nu^{(k)} \\ \boldsymbol{\lambda}_t^{(k)} \end{pmatrix} = \\ = \begin{pmatrix} \mathbf{f} \\ \mathbf{D}_{\mathcal{A}_\nu} \mathbf{d} \\ \mathbf{0} \end{pmatrix}. \end{aligned}$$

- (3) Set $\text{err}^{(k)} := \|\mathbf{u}^{(k)} - \mathbf{u}^{(k-1)}\| / \|\mathbf{u}^{(k)}\|$. If $\text{err}^{(k)} \leq \varepsilon_u$, return $\mathbf{u} := \mathbf{u}^{(k)}$, $\boldsymbol{\lambda}_\nu := \boldsymbol{\lambda}_\nu^{(k)}$, and $\boldsymbol{\lambda}_t := \boldsymbol{\lambda}_t^{(k)}$.

(4) Set $k := k + 1$ and go to step (1).

As we deal in step (2) with non-symmetric matrices with generalized saddle-point structure we use BiCGSTAB.

Symmetric case. The only difference when we want to propose the algorithm for solving the discrete problem with Tresca friction is in the definition of inactive sets when $\mathcal{F}_i(\lambda_{\nu,i} + \rho(\mathbf{Nu} - \mathbf{d})_i)^+$ is substituted by g_i . The algorithm then differs only in the step (2). The matrix now can be made symmetric and we can use CGM. The algorithm stays the same and the step (2) is presented below.

Algorithm *ActiveSetTresca*

(2) Solve:

$$\begin{pmatrix} \mathbf{K} & \mathbf{N}_{\mathcal{A}_\nu}^\top & \mathbf{T}_{\mathcal{A}_t}^\top \\ \mathbf{N}_{\mathcal{A}_\nu} & \mathbf{0} & \mathbf{0} \\ \mathbf{T}_{\mathcal{A}_t} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u}^{(k)} \\ \lambda_{\nu, \mathcal{A}_\nu}^{(k)} \\ \lambda_{t, \mathcal{A}_t}^{(k)} \end{pmatrix} = \begin{pmatrix} \mathbf{f} - \mathbf{T}_{\mathcal{I}_t^+}^\top \mathbf{g}_{\mathcal{I}_t^+} + \mathbf{T}_{\mathcal{I}_t^-}^\top \mathbf{g}_{\mathcal{I}_t^-} \\ \mathbf{d}_{\mathcal{A}_\nu} \\ \mathbf{0} \end{pmatrix},$$

and set $\lambda_{\nu, \mathcal{I}_\nu}^{(k)} = \mathbf{0}$, $\lambda_{t, \mathcal{I}_t^+}^{(k)} = \mathbf{g}_{\mathcal{I}_t^+}$, $\lambda_{t, \mathcal{I}_t^-}^{(k)} = -\mathbf{g}_{\mathcal{I}_t^-}$.

In order to solve the discrete contact problem with Coulomb friction, one can use the algorithm *ActiveSetTresca* in each successive approximation. In order to achieve high computational efficiency, we prefer the inexact implementation in which one iteration of the algorithm *ActiveSetTresca* is performed. In other words, we modify the algorithmic scheme *ActiveSetTresca* so that we take $\mathbf{g} := \mathcal{F} \max\{\lambda_\nu^{(k-1)}, 0\}$ in the beginning of each iteration.

3 Implementation

In this section, we interpret each step (2) of algorithm *ActiveSetTresca* as the minimization of a strictly quadratic objective function in terms of λ , for that one can use the conjugate gradient method. We will arrive at the globally convergent dual implementation of the algorithm; see [3].

We start with notations.

$$\mathcal{A} = \{i \mid i \in \mathcal{A}_\nu\} \cup \{i + m \mid i \in \mathcal{A}_t\} \quad \text{and} \quad \mathcal{I} = \{1, 2, \dots, 2m\} \setminus \mathcal{A}.$$

Denote

$$\mathbf{B} = \begin{pmatrix} \mathbf{N} \\ \mathbf{T} \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} \mathbf{d} \\ \mathbf{0} \end{pmatrix}, \quad \lambda = \begin{pmatrix} \lambda_\nu \\ \lambda_t \end{pmatrix},$$

and

$$q(\lambda) = \frac{1}{2} \lambda^\top \mathbf{A} \lambda - \lambda^\top \mathbf{b},$$

where $\mathbf{A} = \mathbf{BK}^{-1}\mathbf{B}^\top$, $\mathbf{b} = \mathbf{BK}^{-1}\mathbf{f} - \mathbf{c}$. Finally note that the gradient to q at λ reads as

$$\nabla q(\lambda) = \mathbf{A} \lambda - \mathbf{b}.$$

The linear system in step (2) of algorithm *ActiveSetTresca* reads as

$$\begin{pmatrix} \mathbf{K} & \mathbf{B}_A^\top \\ \mathbf{B}_A & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u}^{(k)} \\ \boldsymbol{\lambda}_A^{(k)} \end{pmatrix} = \begin{pmatrix} \widehat{\mathbf{f}} \\ \mathbf{c}_A \end{pmatrix}, \tag{3.8}$$

where $\widehat{\mathbf{f}} = \mathbf{f} - \mathbf{T}_{\mathcal{I}_t^+}^\top \mathbf{g}_{\mathcal{I}_t^+} + \mathbf{T}_{\mathcal{I}_t^-}^\top \mathbf{g}_{\mathcal{I}_t^-}$ and the remaining components of $\boldsymbol{\lambda}^{(k)}$ are given by

$$\boldsymbol{\lambda}_{\nu, \mathcal{I}_\nu}^{(k)} = \mathbf{0}, \quad \boldsymbol{\lambda}_{t, \mathcal{I}_t^+}^{(k)} = \mathbf{g}_{\mathcal{I}_t^+}, \quad \boldsymbol{\lambda}_{t, \mathcal{I}_t^-}^{(k)} = -\mathbf{g}_{\mathcal{I}_t^-}. \tag{3.9}$$

The following lemma interprets the solution.

Lemma 1. (i) Let $\mathcal{A} = \emptyset$. Then $\boldsymbol{\lambda}^{(k)}$ is fully determined by (3.9) and

$$\mathbf{u}^{(k)} = \mathbf{K}^{-1} \widehat{\mathbf{f}} = \mathbf{K}^{-1} (\mathbf{f} - \mathbf{B}^\top \boldsymbol{\lambda}^{(k)}).$$

(ii) Let $\mathcal{A} \neq \emptyset$. Then $\boldsymbol{\lambda}_A^{(k)}$ is the minimizer to the problem

$$\min \frac{1}{2} \boldsymbol{\lambda}_A^\top \mathbf{A}_{\mathcal{A}\mathcal{A}} \boldsymbol{\lambda}_A - \boldsymbol{\lambda}_A^\top \widehat{\mathbf{b}}_A, \tag{3.10}$$

where $\widehat{\mathbf{b}}_A = \mathbf{B}_A \mathbf{K}^{-1} \widehat{\mathbf{f}} - \mathbf{c}_A$. The remaining components of $\boldsymbol{\lambda}^{(k)}$ are given by (3.9) and

$$\mathbf{u}^{(k)} = \mathbf{K}^{-1} (\widehat{\mathbf{f}} - \mathbf{B}_A^\top \boldsymbol{\lambda}_A^{(k)}) = \mathbf{K}^{-1} (\mathbf{f} - \mathbf{B}^\top \boldsymbol{\lambda}^{(k)}).$$

In the next lemma, we show how to define the active and inactive sets without the knowledge of $\mathbf{u}^{(k)}$.

Lemma 2. Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(k)}$. It holds:

$$\begin{aligned} \mathcal{A}_\nu &= \{i \in \mathcal{M} : \lambda_{\nu,i} - \rho \nabla_i q(\boldsymbol{\lambda}) > 0\}, \\ \mathcal{I}_t^+ &= \{i \in \mathcal{M} : \lambda_{t,i} - \rho \nabla_{i+m} q(\boldsymbol{\lambda}) > g_i\}, \\ \mathcal{I}_t^- &= \{i \in \mathcal{M} : \lambda_{t,i} - \rho \nabla_{i+m} q(\boldsymbol{\lambda}) < -g_i\}. \end{aligned}$$

Finally, we show how to determine $\boldsymbol{\lambda}^{(k)}$ via the constrained minimization problem.

Lemma 3. Let $\mathcal{A} \neq \emptyset$. Then $\boldsymbol{\lambda}^{(k)}$ determined in step (2) of algorithm *ActiveSetTresca* is the minimizer of the problem:

$$\min q(\boldsymbol{\lambda}) \tag{3.11}$$

$$\text{subject to } \boldsymbol{\lambda}_{\nu, \mathcal{I}_\nu} = \mathbf{0}, \quad \boldsymbol{\lambda}_{t, \mathcal{I}_t^+} = \mathbf{g}_{\mathcal{I}_t^+}, \quad \boldsymbol{\lambda}_{t, \mathcal{I}_t^-} = -\mathbf{g}_{\mathcal{I}_t^-}. \tag{3.12}$$

Algorithm *ActiveSetTresca* (Dual Version: Implementation 1)

(0) Set $k := 1, \rho > 0, \varepsilon_\lambda > 0, \boldsymbol{\lambda}^{(0)} \in \mathbb{R}^{2m}$.

(1) Define the active and inactive sets at $\lambda = \lambda^{(k-1)}$:

$$\begin{aligned} \mathcal{A}_\nu &= \{i \in \mathcal{M} : \lambda_{\nu,i} - \rho \nabla_i q(\lambda) > 0\}, \\ \mathcal{I}_t^+ &= \{i \in \mathcal{M} : \lambda_{t,i} - \rho \nabla_{i+m} q(\lambda) > g_i\}, \\ \mathcal{I}_t^- &= \{i \in \mathcal{M} : \lambda_{t,i} - \rho \nabla_{i+m} q(\lambda) < -g_i\}, \\ \mathcal{I}_\nu &= \mathcal{M} \setminus \mathcal{A}_\nu, \quad \mathcal{A}_t = \mathcal{M} \setminus (\mathcal{I}_t^+ \cup \mathcal{I}_t^-) \end{aligned}$$

and \mathcal{A}, \mathcal{I} .

(2) If $\mathcal{I} \neq \emptyset$, set $\lambda_{\mathcal{I}}^{(k)}$ so that $\lambda_{\nu, \mathcal{I}_\nu}^{(k)} = \mathbf{0}$, $\lambda_{t, \mathcal{I}_t^+}^{(k)} = \mathbf{g}_{\mathcal{I}_t^+}$, $\lambda_{t, \mathcal{I}_t^-}^{(k)} = -\mathbf{g}_{\mathcal{I}_t^-}$.

If $\mathcal{A} \neq \emptyset$, find

$$\lambda_{\mathcal{A}}^{(k)} := \arg \min \frac{1}{2} \lambda_{\mathcal{A}}^\top \mathbf{A}_{\mathcal{A}\mathcal{A}} \lambda_{\mathcal{A}} - \lambda_{\mathcal{A}}^\top \widehat{\mathbf{b}}_{\mathcal{A}}.$$

(3) Set $err^{(k)} := \|\lambda^{(k)} - \lambda^{(k-1)}\| / \|\lambda^{(k)}\|$. If $err^{(k)} \leq \varepsilon_\lambda$, return

$$\mathbf{u} := \mathbf{K}^{-1}(\mathbf{f} - \mathbf{B}^\top \lambda^{(k)}), \quad \lambda := \lambda^{(k)}$$

and stop.

(4) Set $k := k + 1$ and go to step (1).

The standard result on convergence of the semi-smooth Newton method [5] requires an initial iteration "sufficiently close" to the solution. Unfortunately, it is impractical from the point of view of computations. To overcome this drawback, it is proposed by Jungho Lee in [2] to find the initial iteration by projecting an auxiliary solution of an appropriate unconstrained problem.

Let us briefly comment another *globalization strategy*. Its idea consists in inexact solving of the inner linear systems using several steps of the conjugate gradient method so that all inner iterations connected to one sequence $\{\lambda^{(l)}\}$ give the decreasing sequence $\{q(\lambda^{(l)})\}$. Moreover, we adopt from [1] the criterion for the adaptive inner precision control. This implementation of *ActiveSetTresca* is closely related to the MPRGP algorithm, see [1]. The main difference consists in definitions of the active and inactive sets. Nevertheless, the same convergence result may be proved as for the MPRGP, if the initial iteration $\lambda^{(0)}$ belongs to $\mathbb{R}_+^m \times \mathbf{A}(\mathbf{g})$.

References

1. Z. Dostál. *Optimal Quadratic Programming Algorithms, with Applications to Variational Inequalities*. Springer, 2008, New York, 1st ed.
2. J. Lee. *A Strategy of Finding an Initial Active Set for Inequality Constrained Quadratic Programming Problems* Submitted to Optimization Methods and Software.
3. K. Motyčková. *The Semi-Smooth Newton Method for Solving Contact Problems with Friction*. Ph.D. thesis, in preparation.
4. J. Nocedal, S. J. Wright. *Numerical Optimization*. Springer, 2006, New York, 2nd ed.
5. K. Rádková. *Newton-like Methods for Nonlinear Programming*. Master thesis, UP Olomouc, 2010.

Parallel Hybrid Mesh Multiplication

Aleš Ronovský

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
ales.ronovsky@vsb.cz

Abstract. The paper describes full hybrid mesh multiplication algorithm for Code_Saturne software. It is an extension of the package described in [1]. For an accurate solution in many fields of research, it is necessary to work with very fine meshes. The algorithm implements parallel global refinement of mesh containing any basic types of cells. Hybrid meshes are often used for complex simulations in Earth Sciences such as flood modeling, ground, underground or air pollution, etc. This algorithm works with hexahedras, tetras, prims and pyramids. These problems belong to a class of mathematical problems known as Computational Fluid Dynamics (CFD). The paper describes an algorithm for subdivision of current cells into given number of new cells together with computation of new local and global indices. Using described algorithm, mesh with more than billion cells is obtained that allows more accurate CFD simulations than with the coarser mesh. The effectiveness of implemented algorithm is demonstrated on practical examples and scalability results are also presented.

1 Introduction and Mesh Multiplication

Creating very large meshes of billion cells is a challenge and there are very few open-source parallel mesh generators. Besides, created meshes cannot be easily stored, transferred and for many CFD solvers it is very difficult even to read such meshes. An alternative is to use mesh multiplication(global refinement) firstly mentioned in [2] to create a very large mesh from an initial coarser mesh, already suitable for the CFD solver. This strategy allows to keep stretching between cells and skewness of the initial mesh for the new refined mesh.

We used subdivisions of each of the cells of the original mesh into certain number of cells by using midpoints of its edges and centers of its quadrangle faces to ensure global connectivity of the new vertices and to avoid unnecessary core-to-core communication (Fig. 1). It secures that new coordinates and indices of vertices are computed same way on every single subdomain of whole mesh.

We decided to use Code_Saturne software for CFD simulation and write algorithm for mesh multiplication as a package to that software, because Code_Saturne has been released as open source in 2007 and is distributed under GPL license. Code_Saturne is multi-purpose CFD software, which has been developed by Electricité de France Recherche et Développement since 1997. Several partners and developing teams have been selected to enable petascale capabilities of Code_Saturne and mesh multiplication could be conducive to it.

1.1 Data structure `cs_mesh_t`

To ensure global connectivity of the new elements after refinement and to avoid unnecessary core-to-core communication it is necessary to create global numbering for each element during the preprocessing step of the coarse mesh. Subdivision of cells of the original coarse mesh is done by using midpoints of its edges, its centers of rectangular faces and vertices in center of hexahedral cells.

Since the `Code.Saturne` `mesh_t` structure handles only global numbering for vertices, faces and cells of the mesh and that is not sufficient for the efficient mesh refinement, it was necessary to extend the current mesh storage format by adding a few pieces of information about the mesh. We need to gather the information about the edges and build the cells during pre-processing step of the algorithm. During that step local and global indices of each edge are created. For full hybrid refinement another information is required. Local and global indices for each type of cell and for quadrangle faces are created too. That is required to compute new indices of refined cells and new created vertices during the refinement step of the algorithm.

1. Vertices

- From coarse mesh: keep indices
- Edge vertex: $n_vertices + edge_idx$
- Quadrangle face vertex: $n_vertices + n_edges + face_idx$
- Hexa cell vertex: $n_vertices + n_edges + n_faces + cell_idx$

2. Faces

- Every face refined into 4
- Refined face: $4*(face_idx - 1) + 1:4$
- New face (cell subdivision): $4*n_faces + T*(cell_idx-1) + 1:T$
- T depends on mesh (12 for hexa, tetra, 10 for prism and pyramids)

3. Cells

- New cell: $T*(cell_idx-1) + 1:T$
- T depends on mesh (8 for hexa, tetra, prism and 10 for pyramids)

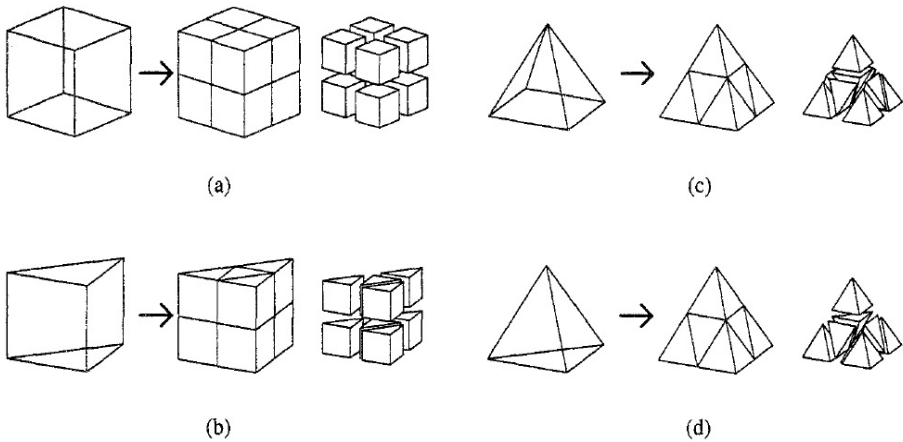


Fig. 1. 3D element refinement. (a) hexahedras, (b) prisms, (c) pyramids, (d) tetrahedras.

Full hybrid mesh refinement algorithm (hexahedral mesh):

- Input: coarse mesh.
- Pre-processing:
 - Create local/global numbering,
 - * Basic: vertices, cells, interior and border faces
 - * Edges
 - * Hybrid: quadrangles, hexahedras, tetrahedras, prisms, pyramids
 - Create face to edge connectivity,
 - Define cells.
- Refinement:
 - Create a new vertex in the middle of each edge,
 - Create a new vertex in the center of each quadrangle face,
 - Refine all the faces,
 - All new faces inherit family and group from parent.
- Cell refinement (of each single cell):
 - Preparation:
 - Create a new vertex in the center of gravity in case of hexahedral cell,
 - Order faces of the cell to ensure positiveness of normal vectors,
 - Prepare indices of vertices,
 - Cell subdivision:
 - Refine the cells,
 - * Refine hexahedras (only hexahedras created)
 - * Refine tetrahedras (only tetrahedras created)
 - * Refine prisms (only prisms created)
 - * Refine pyramids (4 tetrahedras and 6 pyramids created)
 - Create new interior faces,
 - Assign proper face to cell connectivity to each new face and cell
- Output: refined mesh.

2 Parallel Mesh Multiplication for Code_Saturne

The basic capabilities of Code_Saturne enable the handling of either incompressible or expandable flows with or without heat transfer and turbulence. There are some dedicated modules available for specific physics such as radiative heat transfer, combustion (gas, coal, heavy fuel oil, ...), magneto-hydrodynamics, compressible flows, two-phase flows (Euler- Lagrange approach with two-way coupling), or atmospheric flows. Code_Saturne is portable on all Linux and UNIX platforms tested so far (HP-UX, Solaris, Cray, IBM Blue Gene, Tru64, ...). It runs in parallel using MPI on distributed memory machines (clusters, Cray XT, IBM Blue Gene, ...). It is based on a co-located Finite Volume approach that accepts meshes with any type of cell (tetrahedral, hexahedral, prismatic, pyramidal, polyhedral, ...) and any type of grid structure (unstructured, block structured, hybrid, conforming or with hanging nodes, ...).

The process of mesh multiplication is then executed in parallel on distributed parts of the mesh, i.e. each parallel process takes care of its own part of the mesh, but the faces on the subdomain interface are shared with more processors. Because of extended construction of global numbering during preprocessing of the mesh, it is guaranteed that the refined mesh match on subdomain interface for each basic element type.

3 Subdomain mesh refinement

Several mesh partitioning routines are implemented in Code_Saturne, as well as a possibility of using some of standard mesh partitioners like Metis, Zoltan, etc. In parallel environment, ParMetis package can be also used.

The mesh refinement is executed in parallel on distributed parts of the mesh and it is done using local variables. During the final global re-indexation, special care has to be given to faces on the subdomain interfaces that are shared with other processors. Very fine meshes can be obtained applying the implemented mesh multiplication algorithm recursively. If required, the next levels of refinement can be processed in the same way as the first level.

4 Tests and results

Performance and scalability tests of the Parallel Mesh Multiplication algorithm have been carried out on different machines with different architectures, for example HECToR (CRAY XE6), Blue Joule and Mira (both IBM Blue Gene/Q). For most tests, the simulation is stopped just after refinement step, even if runs of 10 time-steps for the 13B cell mesh and 1 time-step for the 106B cell mesh have been performed with Code_Saturne. The pressure is solved by an Algebraic Multigrid algorithm (V-cycle), with the Conjugate Gradient as a solver and the Jacobi algorithm is used for the velocity components.

Test example

The test cases consist of LES in staggered distributed tube bundles (see Fig. 4). The configuration is the one experimentally studied by Simonin and Barcouda [8], and the original mesh is obtained by copying and shifting several elemental patterns made of a whole tube in its centre and four quarter of tubes in the corners. The elemental pattern contains about 13 million cells (2-D cross-section: 100,040 cells; 3rd direction: 128 layers).

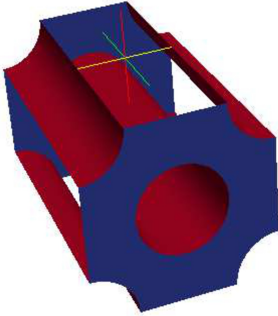


Fig. 2. Single elemental pattern of tubes (13M cells)

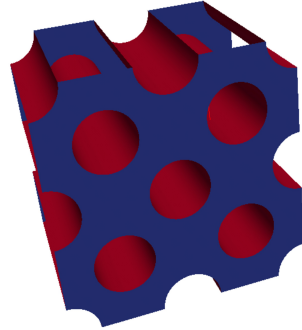


Fig. 3. Four patterns joined (51M cells)

We present an implementation of a multilevel mesh multiplication (global refinement) algorithm for multibillion cell mesh simulations. The mesh multiplication algorithm takes a coarse mesh and subdivides each of the current cells into given numbers of cells, subdivides the boundaries and computes the global indices of the new cells. Three coarse meshes are used, with a single elemental cell (13M cells), with 2 elemental cells joined together (26M cells) and 4 elemental cells joined together (51M cells). Up to 3 levels of refinement are used, but the algorithm is able to generate higher levels of refinement. We obtained meshes of 6.6B, 13B and 26B cells, respectively. Some results from BlueJoule machine are shown in Table 1.

5 Conclusion

The time spent in the algorithm used in Code_Saturne for mesh multiplication has proven to be modest in comparison with the time of whole computation (at most a couple of minutes compared to hours or days). More testing will be performed on other architectures and on different and/or more complex (hybrid) meshes.

Parameters of given mesh	Level of MM:		0	1	2	3	
	no. of	cells		51M	409M	3.3B	26B
		border faces		1.7M	6.7M	27M	108M
		interior faces		153M	1.2B	9.8B	78B
		vertices		52M	413M	3.3B	26B
Number of cores	16k cores	time [s]	-	2.0(4k)	4.13	23.7	
		cells per core	3k	100k	200k	1.6M	
	32k cores	time [s]	-	1.89(8k)	3.5	14.5	
		cells per core	1.5k	50k	100k	800k	
	65k cores	time [s]	-	1.2(16k)	2.79	-	
		cells per core	800	25k	50k	400k	

Table 1. Performance of four tube bundles example on Blue Gene/Q cluster. (51M)

6 Acknowledgements

This work has been supported by the PRACE project under grant agreement no. RI-211528 and FP7-261557. The last part of the work is also supported by HPC Europa2. This work was also supported by projects of major infrastructures for research, development and innovation of Ministry of Education, Youth and Sports(Czech Republic) with reg. num. LM2011033.

References

1. Parallel Mesh Multiplication and its Implementation in Code_Saturne. A. Ronovský, P. Kabelíková, V. Vondrák, C. Moulinec, Civil-Comp Proceedings ISSN 1759-3433. (doi:10.4203/ccp.101.11)
2. P. Kabelikova, A. Ronovsky, V. Vondrak; Parallel Mesh Multiplication for Code_Saturne, PRACE project (FP7/2007-2013)
3. C. Moulinec; M. J. B. M. Pourquié; B. J. Boersma; T. Buchal; F. T. M. Nieuwstadt Direct Numerical Simulation on a Cartesian Mesh of the Flow through a Tube Bundle. International Journal of Computational Fluid Dynamics, Volume 18, Issue 1 January 2004 , pages 1 - 14
4. M.L.Staten and N.L. Jones; Local refinement of three-dimensional finite element meshes, Engineering with Computers (1997), 13: 165- 174
5. Developing Code_Saturne® for Computing at the Petascale, Y. Fournier, J. Bonelle, C. Moulinec, Z. Shang, A.G. Sunderland, J.C. Uribe, in Press for Computers and Fluids Journal, Special Edition 2011.
6. Development of a Two-velocities Hybrid RANS-LES Model and its Application to a Trailing Edge Flow. J. Uribe, N. Jarrin, R. Prosser, D. Laurence, Journal of Flow Turbulence and Combustion (DOI: 10.1007/s10494-010-9263-6), 2010.
7. Y. Fournier, J. Bonelle, C. Moulinec, Z. Shang, A. G. Sunderland and J. C. Uribe. Comp. & Fluids, **45**, Issue 1, (2011), pp. 103-108.
8. O. Simonin and M. Barcouda. 4th Int. Symp. on Apps of Laser Anemometry to Fluid Mechanics. (1988).

Applied graph theory in Markov chains reliability analysis

Pavel Skalny

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
pavel.skalny@vsb.cz

Abstract. The paper presents reliability analysis which was held for an industrial firm. Aim of the paper is to estimate the probability that the firm fails to satisfy an order to its industrial partners. Since the production process consists of many components, Markov chains- describe the issue ideally. The method is suitable for many systems where we can easily distinguish various states. To quantify the production for each state an algorithm inspired by Graph theory is used.

Keywords: reliability; Markov chains, Flow in Network

1 Introduction

The reliability of production plays fundamental role in an industrial sphere. Nowa-days the reliability of industry process is on a high level. It increases by improving the quality of each component or by redundancy of the production process. Even though it is the top reliability process, there is still a chance of system failing or system proceeding limitedly. In our case we analyse the process which has no redundancy. Thus the information about the probability of the systems failures for certain time period t is very valuable. The research presented here, was motivated by the practical problem. Analysed company was asked what the probability of production failure was. Knowledge of risk, that the order wont be delivered in time, is important for the partners firm to establish sufficient goods supplies. Our aim is to solve the problem by applying discrete time Markov chains (from now on we will understand Markov chain as a discrete time model) a well-known method of stochastic modelling. Nowadays memoryless method of Markov chains has various fields of application in physics, statistics, computer science etc. Markov chains were also effectively applied in stochastic and reliability modelling of industry process [2, 3]. The advantage of Markov chains is that calculations of its results are very easy in principle. On the other hand the amount of computations usually in-creases rapidly with amount of states and time. Many numerical algorithms were developed to deal with the problem of high amount of computations [4]. Another way how to deal with the

problem is to utilize simulations method. In this paper we have chosen the second approach, so well known Monte Carlo method was applied. The simulation approach was used because even in our case study the amount of possibilities was too large to be solved analytically.

2 Markov Chains

Markov chain is a random process with a discrete time set $T \subset \mathbb{N} \cup \{0\}$, which satisfies so called Markov property. The Markov property means that the future evolution of the system depends only on the current state of the system and not on its past history.

$$P\{X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n\} = P\{X_{n+1} = x_{n+1} | X_n = x_n\}. \quad (1)$$

where:

X_1, \dots, X_n is a sequence of random variables. The index denotes certain time $t \in T$ x_1, \dots, x_n is a sequence of states in time $t \in T$. As a transition probability p_{ij} we regard probability, that the system changes from the state i to the state j .

$$p_{ij} = P\{X_{n+1} = x_j | X_n = X_i\}. \quad (2)$$

Matrix P , where p_{ij} is placed in row i and column j , is for all admissible i and j called transition probability matrix.

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{pmatrix}. \quad (3)$$

Clearly all elements of the matrix P satisfy the following property:

$$\forall i \in \{1, 2, \dots, m\} : \sum_{j=1}^n p_{ij} = 1. \quad (4)$$

As $v(t)$ we denote vector of probabilities of all the states in the time $t \in T$. As $v(0)$ we denote an initial vector. Usually all its values are equal to zero except the i^{th} , which is equal to 1. It can be proved:

$$v(t) = v(0) \cdot P^t. \quad (5)$$

3 Calculating of input probabilities

In the previous application a reliability analysis of a part of an industry process was realized. Since we distinguish two states-in order 1 or in fail 0 for each machine, the whole system could occur in one of the $2n$ states where n is an amount

of machines of industry process. The analyzed machines were distributed parallel. Thus it was easy to calculate a whole production of each state-the production of the certain state was calculated as a sum of production of functional states. More complicated situation occurs when the system is not distributed parallel. The aim of this paper is to present a solution of the given problem by the flow in network theory. The firm, which our reliability analysis is made for, is a medium size company specialized in thin layer coating of hard materials. The company has requested not to publish its name and not to publish important data of its production. Thus some information (for example production of certain machines) will be demonstrated as a relative variable. Although the production process of the firm is much more sophisticated, only crucial part of the process will be analysed. We will calculate probabilities that the system fails or works under the condition that the rest of the process works properly. The analysed part of a process consists of 4 machines, where each of them could work properly- state 1, or could be in failure -state 0. Since we have 4 machines we can distinguish $2^4 = 16$ different states. The huge data set (over three years) of fails and the length of repairs was granted for each machine. We choose one hour interval as an appropriate period of each time step. The period of one hour was chosen, because it is approximately equal to the length of one coating process. In the first data source, there was an information about the frequency of fails of each machine. The data were obtained from reports about the failure. We expect that the data of fails come from exponential distribution. Thus to estimate probability that the system fails during one hour we calculated the expected value as an average length of period fail- X :

$$p_f = \frac{1}{Q}. \quad (6)$$

In the data about length of periods between two fails, there were few outlier measurements. These values increased the arithmetic mean of the variable Q significantly. After the discussion with the production manager we removed the outlying measurements from the data. In the second data source file, there was information about the length of repair time. Using the maximum likelihood method we estimated the probability p_r which says that a machine will be repaired within one hour:

$$p_r = \frac{\Delta V}{V}. \quad (7)$$

where: V is an amount of all repairs that were realized. ΔV is an amount of all repairs that lasted less than one hour. To calculate the transition probability matrix P we calculated probabilities p_f , p_r for each machine. The calculated probabilities for given machines a, b, c, d are presented in the following table.

In the second step we have to define all of 2^6 states. As a state we will understand any of the situation that all machines work properly.

probability	A_1	A_2	B_1	B_2	C_1	C_2
p_f	0.104	0.0063	0.0048	0.0031	0.004	0.0033
p_r	0.36	0.36	0.4	0.375	0.24	0.19

4 Application

In the previous application a reliability analysis of a part of an industry process was realized. Since we distinguish two states-in order '1' or in fail '0' for each machine, the whole system could occur in one of the 2^n states where is an amount of machines of industry process. The analyzed machines were distributed parallel. Thus it was easy to calculate a whole production of each state-the production of the certain state was calculated as a sum of production of functional states. More complicated situation occurs when the system is not distributed parallel. The aim of this paper is to present a solution of the given problem by the flow in network theory.

In this chapter we will present an example how to use a flow in network theory in a reliability analysis. At the first we will define states of the industry process and calculate a maximum production of each state by using a well known ford fulkerson's algorithm.

The industry process consists of 6 machines. At each machine we distinguish 2 different states 1-work, 0-in fail. Thus the industry process consists of 2^6 different states. As a certain state of the process we will understand an ordered 6-tuple of 1,0. Let us describe the industry process of a firm with an oriented network with exactly one source and one sink. Every machine is represented as a vertex V . The begin and the end of the process are represented as a source and a sink. The begin of a process consists of acceptance of goods and division between machines of the process. The end of a process consists of product inspection and expediting to the customer. Oriented edges describe the direction of the production process. Labelling of edges-capacity represents the maximum amount of goods processed by the certain machine, which aims towards further machine. Since the maximum amount of processed goods is limited by the capacity of industry machines and not by the primeval acceptance end final expedition of the goods, the edges incident with source and sink are labelled by infinity. For each of 2^6 state we will find the maximum flow in a network. For each state the edges incident with the vertices representing the machines occurred in failure will be taken from the network. To calculate the maximal production of each state a max flow in network we will use a well known Ford Fulkerson algorithm For searching the graph edges a "Breadth-first" search was used.

5 Conclusion

In this paper we presented primeval reliability analysis of the industry firm. The analysis will be also processed as a report for the management of the firm. The probabilities of the failure will be also calculated for longer periods. We can

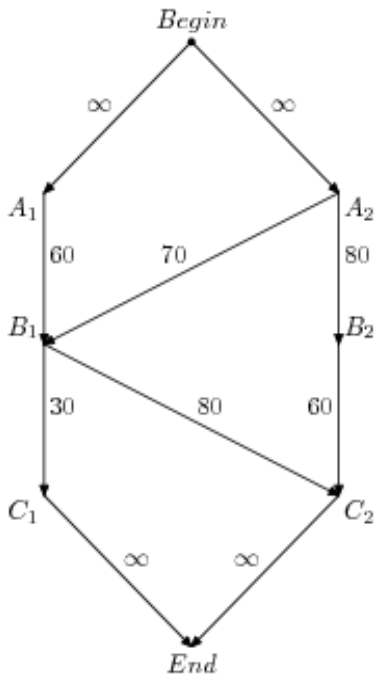


Fig. 1. Structure of the process

conclude, that a flow in network theory seems to be appropriate in solving the non parallel distributed systems.

References

1. Skalny P., Krajc B., *Discrete-Time Markov Chains in Reliability Analysis-Case Study*, Soco 2012
2. Moura M. C. *Semi Markov decision process for determining multiobjective optimal condition-based replacement policies*, ESREL, 2000.
3. Koutras M. V. *On a Markov chain approach for the study of reliability structures*, *Journal of Applied Probability*, Journal of Applied Probability, 1999.
4. Stewart W. *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, 1994.

Comparison of numerical codes for diagonalization of Hermitian matrices

Martin Stachoň

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
`martin.stachon@vsb.cz`

Abstract. Diagonalization of matrices is an important numerical task, especially in quantum chemistry, structural analysis, control theory and other areas. In this survey, we present an overview of algorithms for carrying out the diagonalization, and available numerical libraries implementing these algorithms. Benchmarks are presented on a problem from semiclassical molecular dynamics, comparing the performance of different algorithms and implementations.

Keywords: Hermitian matrix, diagonalization, numerical library

1 Algorithms

In this article, we discuss algorithms for diagonalization of dense, Hermitian matrices, which equals to finding all eigenvalues and corresponding eigenvectors. A whole class of algorithms is at hand, when only several eigenvalues of a large, sparse matrix are desired, such as the Lanczos method. These algorithms are not discussed here, refer eg. to [3].

The algorithms available for diagonalization of a general Hermitian matrix, are, in historical order: the Jacobi method, Householder tridiagonalization followed with QR method or Divide and Conquer, the newest algorithm is Dhillon's MRRR.

1.1 Jacobi method

Named after Carl Gustav Jacob Jacobi, who published it in 1846. The algorithm transforms a matrix to diagonal one using a series of orthogonal transformations, which preserve eigenvalues. The transformation matrices are rotation matrices in the form of:

$$G(p, q, \phi) = \begin{matrix} & & p & & q & & \\ & & & & & & \\ & & & & & & \\ p & \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos \phi & \cdots & -\sin \phi & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ q & \begin{pmatrix} 0 & \cdots & \sin \phi & \cdots & \cos \phi & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \end{pmatrix} \end{matrix}$$

The variable ϕ can be chosen so that matrix element p, q is zeroed out after applying the transformation. The algorithm is an iterative process, where the largest element in absolute value is zeroed out in each iteration. The disadvantage of this method is that a once zeroed element can become non-zero again in subsequent iteration. Thus the algorithm iterates until the largest element is above a given tolerance limit.

The advantages of this algorithm is that it is easy to implement and has high accuracy [5]. However, it cannot compete with newer methods performance-wise.

1.2 Tridiagonalization

Matrix tridiagonalization is used as a first step for the more advanced algorithms. The basic idea is the same as in Jacobi algorithm. However, the matrix is transformed to a tridiagonal one, and if the order of transformation is correctly chosen, tridiagonalization is done in a finite number of steps. There are two ways to perform tridiagonalization: Givens and Householder's. Givens method uses the same rotation transformation matrices as Jacobi, while Householder method uses reflection transformation matrices in the following form:

$$P = I - 2vv^T, v^T v = 1$$

Householder method requires half the number of floating point operations, and so it is the one used in current numerical libraries. After the matrix is made tridiagonal, QR, Divide and Conquer or MRRR algorithm is used for the diagonalization.

1.3 QR method

is an iterative algorithm based on QR matrix factorizations, developed in the 60's. It is usually used on a tridiagonal matrix. It can be shown that the sequence $R_{k-1}Q_{k-1}$, where $Q_{k-1}R_{k-1}$ is QR factorization of the previous matrix in the sequence, converges to a upper triangular matrix with eigenvalues on the diagonal. QR factorization can be computed using Gram-Schmidt process, or using Householder transformations.

1.4 Divide and Conquer

is a new method, which emerged in the 90's. The idea is that a tridiagonal matrix is almost block diagonal:

$$T = \left(\begin{array}{c|c} T_1 & \beta \\ \hline \beta & T_2 \end{array} \right) = \left(\begin{array}{c|c} \widehat{T}_1 & \\ \hline & \widehat{T}_2 \end{array} \right) + \left(\begin{array}{c|c} & \beta \\ \hline \beta & \beta \end{array} \right)$$

The diagonalization of the blocks can be performed in parallel (this is the divide step), using QR algorithm or by recursive Divide and Conquer. The next stage of the algorithm is to conquer - find diagonalization of the original matrix, which leads to a secular equation, which can be solved using Newton's method.

1.5 MRRR

Multiple Relatively Robust Representations is the newest and fastest algorithm, published by Dhillon in his Ph.D thesis in 1997. Unlike QR, MRRR is a $O(n^2)$ algorithm. MRRR can be seen as a sophisticated version of Inverse Iteration. It works with $LDL^T - \sigma I$ factorization of the tridiagonal matrix, called Relative Robust Representation, because a small changes in the nontrivial entries of L, D cause small change in small eigenvalues of LDL^T . A representation tree is built so that each eigenvalue is separated from others. Eigenvectors are computed using one step of inverse iteration and eigenvalues are refined using either bisection or dqds algorithm. For a more detailed description, see [6].

2 Numerical libraries

We have chosen the most popular numerical libraries to perform benchmarks with. Other interesting libraries not included in this benchmark are PLASMA, MAGMA, ELPA, SLEPc and ARPACK, these will be considered in a future work.

2.1 LAPACK

LAPACK is a dense linear algebra package. It is a descendant of classical LINPACK and EISPACK packages. Concerning the symmetric eigenvalue problem, solution is done in three steps:

1. Tridiagonal reduction of matrix using Householder transformations, using routine xSYTRD. Transformations are done with BLAS level 3 routine xSYR2K, which accounts for at most half of the computational effort. The reduction in total requires $4/3n^3 + O(n^2)$ flops.

2. For solving the tridiagonal problem, there are three options:
 - (a) xSTERF, the Pal-Walker-Kahan variant of the QL or QR algorithm. It uses only scalar floating-point operations, without scope for BLAS 2 or 3.
 - (b) xSTEDC introduced in LAPACK 2.0, a divide-and-conquer algorithm. It can exploit BLAS 2 or 3 and requires fewer flops.
 - (c) xSTEGR introduced in LAPACK 3.0, the MRRR algorithm
3. Backtransformation of eigenvectors of the tridiagonal matrix to the original matrix.

2.2 Intel MKL

The Math Kernel Library consists of BLAS, LAPACK and other numerical packages optimized by Intel. The algorithms are the same as in LAPACK.

2.3 BLAS

BLAS is a low-level numerical package, providing vector operations (BLAS 1), matrix-vector operations (BLAS 2) and matrix-matrix operations (BLAS 3). LAPACK and other packages rely heavily on BLAS. The calling interface is standardized, so different BLAS implementations can be linked with LAPACK. As shown in the benchmarks, an optimized BLAS library can drastically improve LAPACK performance. We have used several BLAS implementations:

- Reference BLAS implementation, without optimizations.
- ATLAS, an automatically tuned BLAS (and partially LAPACK). It uses sophisticated methods to produce optimized BLAS for target platform at compile time.
- GotoBLAS2, written by Kazushige Goto, uses heavily optimized hand-written assembly routines.

3 Benchmarks

3.1 Problem

As test matrices, we used electronic Hamiltonian matrices generated using software MULTIDYN [1]. The matrix is constructed using the *diatomics-in-molecules* method with included spin-orbit coupling for a system of 100 or 1000 atoms of Argon atoms in linear geometry, spaced 15 Å apart. The dimension of the matrix is $6N$, where N is the number of atoms.

3.2 Methodology

We have benchmarked the time required to perform diagonalization with LAPACK (using reference BLAS, GotoBLAS and ATLAS) and MKL using the three main methods - QR, DC (Divide and Conquer) and MRRR. The computer had a AMD Opteron 8380 processor, we were using a single core. It should be noted that the performance of MKL might not be optimal as it is designed for Intel processors.

Table 1. Computation times in seconds for diagonalization of Hamiltonian matrix for 100 and 1000 atoms.

Library	$N = 100$	$N = 1000$
LAPACK Reference BLAS QR	1,79	3359,61
LAPACK Reference BLAS DC	1,30	2970,16
LAPACK Reference BLAS MRRR	1,44	2161,22
LAPACK GotoBLAS QR	1,21	2291,83
LAPACK GotoBLAS DC	0,70	1549,43
LAPACK GotoBLAS MRRR	0,72	544,36
LAPACK ATLAS QR	1,50	2631,07
LAPACK ATLAS DC	1,78	2026,35
LAPACK ATLAS MRRR	0,99	891,92
MKL QR	1,17	1287,23
MKL DC	0,63	740,89
MKL MRRR	0,89	773,44

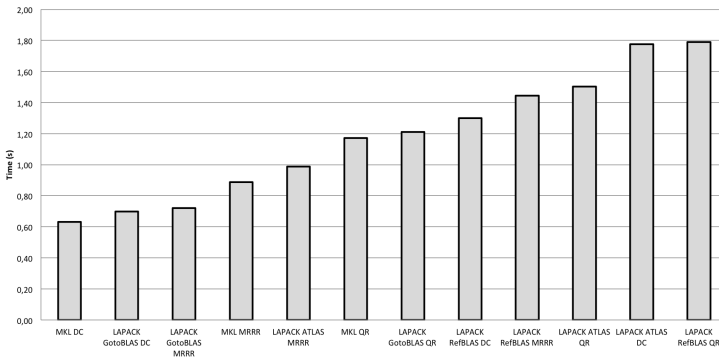


Fig. 1. Computation times for diagonalization of Hamiltonian matrix for 100 atoms.

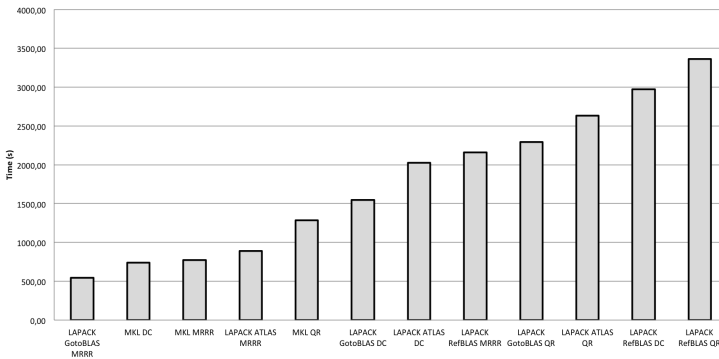


Fig. 2. Computation times for diagonalization of Hamiltonian matrix for 1000 atoms.

4 Conclusions

The benchmark shows an important fact - which can be applied generally to all numerical software: it is important to have an efficient implementation of the low-level code - which comes down to knowing the guts of the CPU (cache sizes and alignment, vector instructions etc.). The speed ratio between a naive BLAS implementation and a heavily optimized one (GotoBLAS, ATLAS or MKL) is of factor 1.5 - 4. However, the algorithms themselves are also important. While at smaller matrix sizes, the differences are small, but at larger matrix sizes the better asymptotic complexity of the new MRRR algorithm begins to shine. The overall speed improvement, when comparing QR algorithm with reference BLAS vs. MRRR algorithm with hand-optimized BLAS is of factor 6 for the larger matrix. This knowledge will be used in development of MULTIDYN software for non-adiabatic molecular dynamics, where matrix diagonalization is the most time consuming task.

5 Acknowledgements

Financial support by IT4Innovations Centre of Excellence project, reg.no. CZ.1.05/1.1.00/02.0070 supported by Operational Programme 'Research and Development for Innovations' funded by Structural Funds of the European Union and state budget of the Czech Republic and Grant of SGS No. SP2013/191.

References

1. Multiple authors, MULTIDYN software, <http://moldyn.vsb.cz/multidyn> (1 July 2013)
2. D. Hrivnak, R. Kalus, F. X. Gadea, Effect of the spin-orbit coupling in the dissociation dynamics of rare-gas ionic trimers, *Europhys. Lett.*, 71 (1), pp. 42–48 (2005)
3. Y. Saad : *Numerical Methods for Large Eigenvalue Problems*, Second edition (2011)
4. G. H. Golub, C. F. Van Loan, *Matrix Computations*, Third edition, Johns Hopkins University Press (1996)
5. J. Demmel, K. Veselic, Jacobi's method is more accurate than QR, <http://www.netlib.org/lapack/lawnspdf/lawn15.pdf> (11 May 2013)
6. I. S. Dhillon, *A New $O(N^2)$ Algorithm for the Symmetric Tridiagonal Eigenvalue/Eigenvector Problem*, PhD. Thesis, University of California, Berkeley (1997).

Elastic image registration with local refinement

Alena Vasatova and Aleš Ronovský

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
alena.vasatova@vsb.cz

Abstract. Image registration arises whenever we need to extract information from different images. Here we present elastic registration method focused on medical usage. Elastic registration is based on physical motivation and supposes that the images are two different observations of elastic body. We can apply any of many methods used in mechanics to solve elasticity (in our approach TFETI). Because medical images usually contain large area of background, thus with better precision we also refine a lot of unnecessary space (especially in 3D). To avoid this we can take coarser grid with local refinement based on image foreground.

Keywords: elastic image registration, local refinement

1 Introduction

Image registration is a crucial step of image processing if there is a need to compare or integrate information from two (or more) images, a reference R and template T . The main task is to find an optimal transformation such that T becomes, in a certain sense, similar to R . These images usually show the same scene, but taken at different times, from different viewpoints or by different sensors.

Image registration is used in various areas. In medical applications it serves to obtain more complete information about the patient, for example to monitor tumor growth, to verify treatment or to compare the patient's data with anatomical atlases [1].

Elastic potential of deformation in connection with image registration has been introduced by Broit [2]. His method can register images with local non-rigid geometric differences. Broit use it to automatically find an optimal mapping between a CT image and an atlas of brain anatomy. The external forces have been derived by correlating intensity-based properties in local regions in the source and target image.

Images in this paper are from department of Oncology at the University Hospital of Ostrava. The images show the cuts in the chest inhale and exhale obtained by CT method.

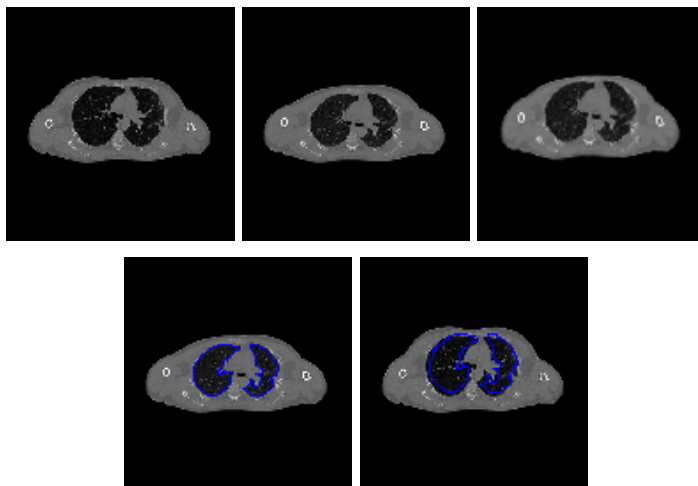


Fig. 1. Example of image registration: image R , image T , transformed image T_φ , image T with segmentation and image R with (transformed) segmentation.

2 Elastic image registration

Elastic registration is based on physical motivation that the images are two different observations of an elastic body, one before and one after a deformation. The transformation $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is splitted into the trivial identity part and the displacement $u : \mathbb{R}^2 \rightarrow \mathbb{R}^2$

$$\varphi(x) = x - u(x). \tag{1}$$

As the regularizer we use linearized elastic potential

$$\mathcal{P}[u] = \int_{\Omega} \frac{\mu}{4} \sum_{j=1}^2 \sum_{k=1}^2 (\partial_{x_j} u_k + \partial_{x_k} u_j)^2 + \frac{\lambda}{2} (\operatorname{div} u)^2 dx, \tag{2}$$

where λ and μ are the so-called Lamé constants. The regularizer has the meaning of internal forces which implicitly constrain the displacement to obey a smoothness criteria. Disadvantage of this linear model is that it assumes small deformations. For larger deformations it can be replace by the viscous fluid model [3].

The optimal transformation is estimated by minimizing the cost function, called distance measure \mathcal{D} , which determines how much is image R , in a certain sense, similar to image T . In this model, it also represents external force, so it pushes the deformable template into the direction of the reference. We choose so-called sum of squared differences (SSD)

$$\mathcal{D}[R, T; u] := \frac{1}{2} \|T_u - R\|_{L_2(\Omega)}^2, \tag{3}$$

where $T_u(x) := T(x - u(x))$. And the forces $f : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ are derived from its Gâteaux-derivative

$$f(x, u(x)) := (R(x) - T_u(x))\nabla T_u(x). \quad (4)$$

The images are represented by the compactly supported mappings $R(x), T(x) : \Omega \rightarrow \mathbb{R}$, where $\Omega = (0, 1)^2$. Hence, $T(x)$ and $R(x)$ denotes the intensities of images at the spatial position x , we set $R(x) = 0$ and $T(x) = 0$ for all $x \notin \Omega$.

If the images are monomodal, i.e. the intensities of corresponding pixels are supposed to be identical, SSD is a reasonable measure for applications. But if the images are multimodal SSD then probably fail. More suitable choices are the mutual information (MI) or the normalized gradient field (NGF).

The partial differential operator associated with the Gâteaux-derivative of the elastic potential is the well-known Navier-Lamé operator. The displacement of the elastic body is then obtained from solution of partial differential equation

$$\mu\Delta u + (\lambda + \mu)\nabla\operatorname{div}u = -f. \quad (5)$$

3 Adaptive mesh discretisation

Uniform mesh (triangle meshes in 2D, tetrahedral meshes in 3D), fine enough to guarantee the accuracy of the solution on whole domain, may cause expensive computation. But medical images contain large area of background, where coarse mesh is sufficient. Mesh adaptation is an effective compromise, providing high accuracy for given region of interest [5].

For computation we use conformal mesh, so we need to deal with a problem between two zones with different levels of refinement. This transition between different levels of refinement occurs on non-interesting part in case of medical images, so we do not need to concern much about mesh quality there.

3.1 Process of adaptive refinement

We start with coarse mesh. For each level of refinement we select elements to refine. For minimizing problem when the pending nodes occurs in the transition zone between two levels of refinement, elements sharing a node with at least one selected element are fully refined too. This nodal neighbours full refinement ensures that difference could not differ more than one level of refinement. All the elements selected to refine are split by midsection.

Elements sharing edge with nodal neighbours are selected as transition elements and ensure conformity of the mesh. They are split using appropriate pattern and then added to final mesh. The rest of the elements are added to final mesh.

Elements to refine are selected on base of binary image (for example see fig. 3) – their intersection with region of interest. This binary image defines foreground and background (can be done automatically by image processing methods). It can be also made by user choice of region of interest, thus the refined region can be even smaller (but interaction is needed). Or selection doesn't have to be based on image, for example error estimation.

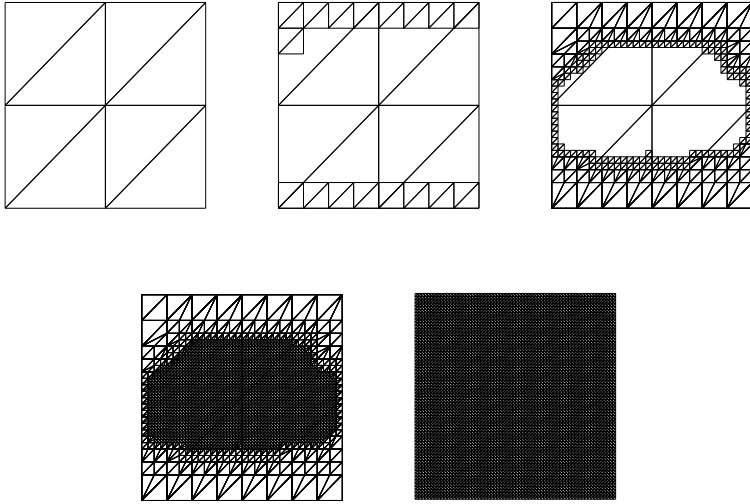


Fig. 2. Example of mesh construction: initial coarse mesh, in progress, final mesh (2 056 nodes) and full mesh (4 225 nodes) for comparison.

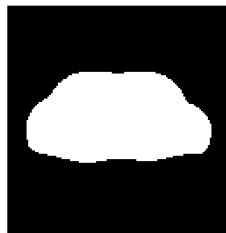


Fig. 3. Binary image based on image R and intensity threshold.

3.2 Transition elements

Single pending node is connected to the opposite vertex, two are connected together and longer is then connected to opposite vertex, see fig. 4. Refined triangles are not "nice" (homothetic to parent), that could modify mesh quality, but it may occurs only outside the region of interest. In 3D, single pending node is connected to the opposite vertex, two are connected together and longer is then connected to opposite vertex.

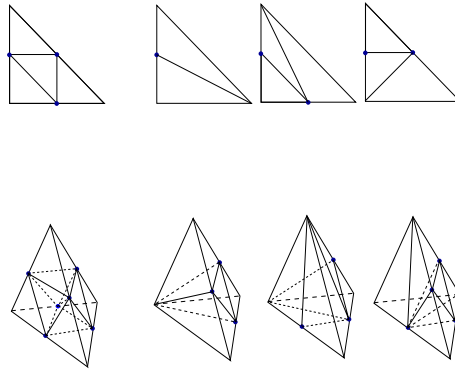


Fig. 4. Transitions elements for 2D and example transitions elements for 3D.

4 Problem Solution

To solve the partial differential equation TFETI [4], a variant of FETI, was used. This method is based on the decomposition of the spatial domain into non-overlapping subdomains, Lagrange multipliers "glued" this subdomains and also enforce the Dirichlet boundary conditions. Thus, all the subdomains are floating and their stiffness matrices will have a priori known kernels – bases of rigid body motion. For more information about the solution itself, see [6].

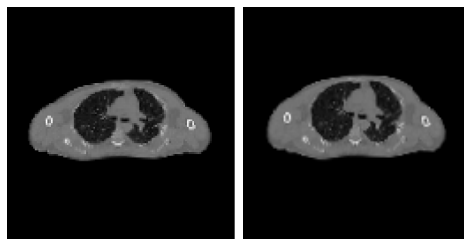


Fig. 5. Solution on full mesh and solution on adaptive mesh.

5 Conclusion

This paper presents elastic image registration solved by TFETI using mesh adaptation. Elastic registration is based on physical motivation of deforming body, thus it fits to register images with local non-rigid differences. We use SSD distance measure for deriving external forces from intensities of monomodal image. Mesh with local refinement exploits the fact that medical images usually contain large area of background where is no need of high accuracy. Navier–Lamé equation are solved by TFETI, a variant of the FETI method for parallel numerical solution of elliptic PDE.

References

1. Modersitzki J. *Numerical Methods for Image Registration*. Oxford : Oxford University Press, 2004. ISBN 0198528418, 9780198528418.
2. Broit C. *Optimal Registration of Deformed Images*, PhD thesis, University Pennsylvania, 1981.
3. Christensen G. E., *Deformable Shape Models for Anatomy*, PhD thesis, Washington University, 1994.
4. Dostal Z., Horak D., Kucera R., Total feti – an easier implementable variant of the feti method for numerical solution of elliptic pde, *Communications in Numerical Methods in Engineering*, Vol.22, No.12, 2006, pp. 1155–1162. ISSN: 1099–0887.
5. G. Nicolas, T. Fouquet, Adaptive mesh refinement for conformal hexahedral meshes, *Finite Elements in Analysis and Design* Vol.67, No.11, 2013, pp. 1–12. ISSN: 0168–874X.
6. A. Vasatova, Elastic image registration with rigid parts solved by TFETI, *WOFEX 2012 proceedings of the 10th annual workshop* , 2012, pp. 441–446. ISBN 978–80–248–2769–8.

Shape Optimization Based on the Shape Calculus and the Boundary Element Method

Jan Zapletal

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
jan.zapletal@vsb.cz

Abstract. In the paper we describe the application of the shape calculus pioneered by Delfour, Hadamard, Sokolowski, Zolésio, and others to industry-inspired 3D shape optimization problems. As the shape of a domain is given by its surface only and the shape gradient is only evaluated on the boundary, the boundary element method provides a useful tool for solving the related state and adjoint boundary value problems. Since shape optimization is an iterative process and in every iteration one has to solve both state and adjoint boundary value problems, the dimension reduction (from a 3D domain to a surface parametrized in 2D) is of great importance.

1 Introduction

Shape optimization plays an important role in many industrial areas. The commonly used approach is based on modifying a couple of parameters describing the given device, e.g., its diameter, curvature, etc. This method is computationally cheap, but limited, as the optimal shape might be totally different from the original design and the handful of parameters might not be able to describe all the necessary changes. Such approach is thus heavily dependent on the experience of the designer who decides on the choice of the design parameters.

The method discussed in this paper is much more general, as it allows to understand every node of the discretized surface as the design variable. Thus, the shape changes can be almost arbitrary and the possibility of finding the optimal shape is higher. In connection with the first-order shape calculus we use the boundary element method, which serves as a great tool for solving such problems.

The paper is divided as follows. In Sect. 2 we describe the reference problem that will be solved using the shape calculus approach described in Sect. 3. The boundary element method is briefly discussed in Sect. 4 and in the final Sect. 5 we provide some numerical examples validating the method.

2 Reference Problem

In the paper we deal with shape optimization based on tracking the Neumann data (see, e.g., [3, 5]) of the electrostatic potential u solving the Dirichlet bound-

ary value problem

$$\begin{cases} -\Delta u = 0 & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_0, \\ u = 1 & \text{on } \Gamma_f. \end{cases} \quad (2.1)$$

For simplicity we only consider a special case with $\Omega \subset \mathbb{R}^3$ denoting a doubly-connected domain, whose boundary is divided into two disconnected components, namely, Γ_0 denoting the fixed part of the boundary, and Γ_f being subject to the optimization process (see Fig. 2.1 for a 2D illustration).

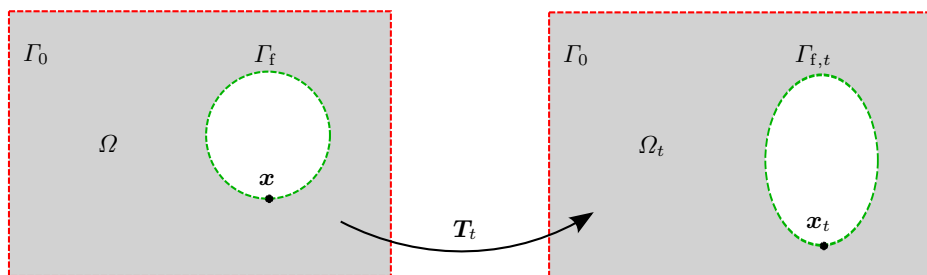


Fig. 2.1. Topology of Ω and its transformation.

The objective is to find the optimal shape of Γ_f minimizing the cost functional

$$J(\Omega, u) := \frac{1}{2} \left\| \frac{\partial u}{\partial \mathbf{n}} - Q \right\|_{L^2(\Gamma_f)}^2 = \frac{1}{2} \int_{\Gamma_f} \left(\frac{\partial u}{\partial \mathbf{n}}(\mathbf{x}) - Q \right)^2 \, d\mathbf{s}_{\mathbf{x}} \quad (2.2)$$

with $0 \leq Q \in \mathbb{R}$. Note that in the electrical engineering context the quantity $\frac{\partial u}{\partial \mathbf{n}}$ can be understood as the normal component of the electric field ∇u on the surface and its minimization is crucial to avoid breakdowns.

3 Shape Calculus

To find the optimal shape we exploit the first-order shape calculus approach (see, e.g., [1, 4, 11]). To describe geometry transformations we define a family of mappings $\mathbf{T}_t: \Omega \rightarrow \mathbb{R}^3$ given as the perturbation of identity, i.e., $\mathbf{T}_t := \mathbf{I} + t\mathbf{V}$ with the artificial time parameter $t \in [0, \tau)$ and the speed field $\mathbf{V}: \Omega \rightarrow \mathbb{R}^3$. The problem now reduces to finding an optimal speed field \mathbf{V} , such that the new iteration $\Omega_t := \mathbf{T}_t(\Omega)$ (see Fig. 2.1) reduces the cost, i.e., the value $J(\Omega_t, u_t)$, where u_t solves the Dirichlet boundary value problem (2.1) stated in the new domain Ω_t .

To minimize the cost functional (2.2) we use the gradient information obtained by the first-order shape calculus. This data can be obtained by taking the Gâteaux derivative in the direction of the speed field \mathbf{V}

$$J'(\Omega, u)(\mathbf{V}) := \left. \frac{d}{dt} \right|_{t=0} J(\Omega_t, u_t) = \lim_{t \rightarrow 0^+} \frac{J(\Omega_t, u_t) - J(\Omega, u)}{t}.$$

To construct an efficient optimization algorithm it is crucial to derive the so-called Hadamard-Zolésio form of the shape derivative (see, e.g., [1, 11]), i.e.,

$$J'(\Omega, u)(\mathbf{V}) = \int_{\Gamma_f} g(\mathbf{x}) \langle \mathbf{V}(\mathbf{x}), \mathbf{n}(\mathbf{x}) \rangle d\mathbf{s}_x. \tag{3.1}$$

Note that the kernel function g appearing in (3.1) does not depend on the speed field \mathbf{V} . Moreover, now we only have to define the optimal mapping \mathbf{V} on the free part of the boundary Γ_f , i.e., $\mathbf{V}: \Gamma_f \rightarrow \mathbb{R}^3$. This fact is quite natural, since defining a new shape of Γ_f already defines the new iteration Ω_t .

For the cost functional (2.2) the Hadamard-Zolésio form of the shape derivative is given by (3.1) with the kernel

$$g(\mathbf{x}) := -\frac{\partial p}{\partial \mathbf{n}}(\mathbf{x}) \frac{\partial u}{\partial \mathbf{n}}(\mathbf{x}) - \frac{H(\mathbf{x})}{2} \left(\left(\frac{\partial u}{\partial \mathbf{n}}(\mathbf{x}) \right)^2 - Q^2 \right) \tag{3.2}$$

with H denoting the additive curvature, u solving (2.1), and p being the solution to the adjoint boundary value problem

$$\begin{cases} -\Delta p = 0 & \text{in } \Omega, \\ p = 0 & \text{on } \Gamma_0, \\ p = \frac{\partial u}{\partial \mathbf{n}} - Q & \text{on } \Gamma_f. \end{cases} \tag{3.3}$$

For a detailed derivation see, e.g., [3, 5].

The most straightforward way to obtain a speed field ensuring decrease in the cost is to define $\mathbf{V} := -g\mathbf{n}$ with the kernel function g from (3.2) and \mathbf{n} denoting the unit exterior normal vector to Ω on Γ_f . Indeed, substituting \mathbf{V} into (3.1) we obtain

$$J'(\Omega, u)(\mathbf{V}) = \int_{\Gamma_f} g(\mathbf{x}) \langle \mathbf{V}(\mathbf{x}), \mathbf{n}(\mathbf{x}) \rangle d\mathbf{s}_x = - \int_{\Gamma_f} g^2(\mathbf{x}) d\mathbf{s}_x \leq 0.$$

Using this approach, we can define the decrease direction for every node of the discretized surface Γ_f . To do this, we have to compute the Neumann data $\frac{\partial u}{\partial \mathbf{n}}$ and $\frac{\partial p}{\partial \mathbf{n}}$. These can be obtained using the boundary element method described briefly in the following section. The curvature H appearing in the kernel is a property of the surface and on triangular meshes can be approximated as in, e.g., [7].

4 Boundary Element Method

To compute the missing Neumann data $\gamma^1 u := \frac{\partial u}{\partial \mathbf{n}}$, $\gamma^1 p := \frac{\partial p}{\partial \mathbf{n}}$ we use the single-layer boundary integral equation

$$(V\gamma^1 u)(\mathbf{x}) = \frac{1}{2} \gamma^0 u(\mathbf{x}) + (K\gamma^0 u)(\mathbf{x}) \quad \text{for } \mathbf{x} \in \partial\Omega \tag{4.1}$$

with the trace operators

$$\begin{aligned} \gamma^0 : H^1_{\Delta}(\Omega) &\rightarrow H^{1/2}(\partial\Omega), & \gamma^0 u &= u|_{\partial\Omega} \text{ for } u \in C^\infty(\overline{\Omega}), \\ \gamma^1 : H^1_{\Delta}(\Omega) &\rightarrow H^{-1/2}(\partial\Omega), & \gamma^1 u &= \frac{\partial u}{\partial \mathbf{n}} \text{ for } u \in C^\infty(\overline{\Omega}), \end{aligned}$$

and the boundary integral operators

$$\begin{aligned} V : H^{-1/2}(\partial\Omega) &\rightarrow H^{1/2}(\partial\Omega), \\ K : H^{1/2}(\partial\Omega) &\rightarrow H^{1/2}(\partial\Omega) \end{aligned}$$

given by

$$\begin{aligned} (V\gamma^1 u)(\mathbf{x}) &:= \int_{\partial\Omega} v(\mathbf{x}, \mathbf{y}) \gamma^1 u(\mathbf{y}) \, d\mathbf{s}_{\mathbf{y}} && \text{for } \mathbf{x} \in \partial\Omega, \\ (K\gamma^0 u)(\mathbf{x}) &:= \int_{\partial\Omega} \frac{\partial v}{\partial \mathbf{n}_{\mathbf{y}}}(\mathbf{x}, \mathbf{y}) \gamma^0 u(\mathbf{y}) \, d\mathbf{s}_{\mathbf{y}} && \text{for } \mathbf{x} \in \partial\Omega. \end{aligned}$$

with v denoting the fundamental solution to the Laplace equation in 3D

$$v(\mathbf{x}, \mathbf{y}) := \frac{1}{4\pi} \frac{1}{\|\mathbf{x} - \mathbf{y}\|}.$$

Using the Galerkin discretization with piecewise constant ansatz and test functions (denoted by ψ_i, ψ_j , respectively) we obtain the system of linear equations

$$\sum_{i=1}^E t_i \langle V\psi_i, \psi_j \rangle_{\partial\Omega} = \sum_{i=1}^E s_i \left\langle \left(\frac{1}{2}I + K \right) \psi_i, \psi_j \right\rangle_{\partial\Omega} \quad \text{for all } j \in \{1, \dots, E\}. \quad (4.2)$$

Note that both boundary value problems (2.1) and (3.3) are of the same type, i.e., the matrices defined by the bilinear forms in (4.2) can be used for solving both problems. Properties of the boundary integral operators ensure unique solvability of both the equation (4.1) and its discretized counterpart (4.2). For more details on the boundary element method consult, e.g., [8, 10, 12–14]

5 Numerical Examples

In the last section we provide numerical experiments validating the approach detailed above. Note that except for the methods described (i.e., the shape calculus and the boundary element method) we also use the multiresolution analysis implemented in the `openFTL` library provided by the University of Cambridge (corresponding papers have not been published yet) and the fast multipole method (see, e.g., [2, 9]) implemented in the `gobem` library provided by TU Graz to speed up the boundary element analysis. See the Acknowledgements section below for more details on this cooperation.

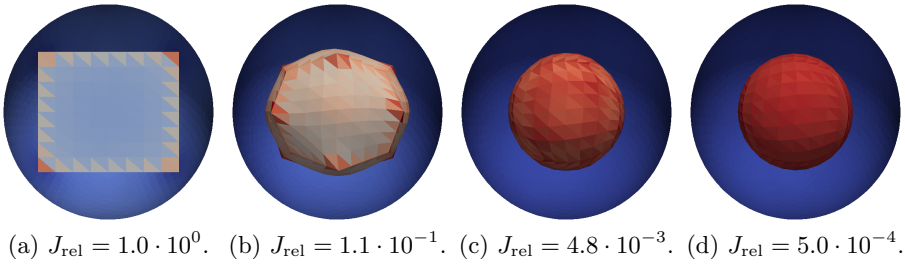


Fig. 5.1. Example 1: initial shape and three optimization stages.

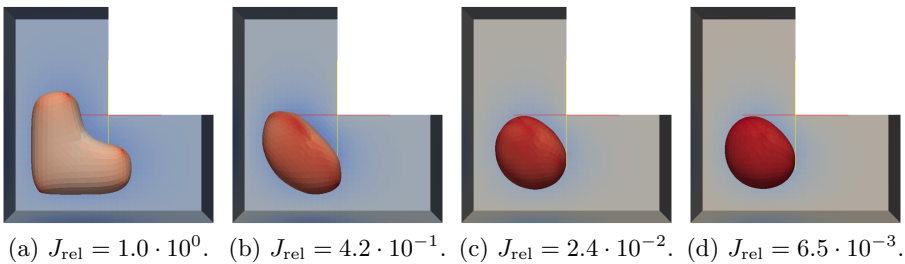


Fig. 5.2. Example 2: initial shape and three optimization stages.

As the first example we consider the situation depicted in Fig. 5.1a, where Γ_0 is given by a unit sphere and the initial shape of Γ_f is given by a box surface placed inside (note that in Fig. 5.1 the sphere is cut, so that the interior surface is visible). We minimize the cost functional (2.2) with $Q = 20$. For this example, the analytic solution is known and the optimal shape of Γ_f is given by a sphere with half the diameter of the exterior one. In Fig. 5.1 we present three stages of the optimization process and the normalized value of the cost functional (2.2), i.e., the initial value is scaled to 1.0. During the optimization process the cost decreases to be finally reduced by approx. 99.95%. Moreover, the final shape of Γ_f is very close to the optimal analytical solution.

The starting geometry for the second example is depicted in Fig. 5.2a, Q is set to 30. In Fig. 5.2 we provide three stages of the optimization process and the normalized values of the cost. Again, the cost is finally reduced by approx. 99.35%. Note that in this case the analytical solution is not known.

Acknowledgements

Most of the work presented in this paper was performed during the author's stay at the Institute of Computational Mathematics at TU Graz from September 2012 to March 2013. The research was supported by the FP7 Marie Curie IAPP Project CASOPT (Controlled Component and Assembly Level Optimisation of Industrial Devices).

Together with the colleagues from TU Graz (G. Of, O. Steinbach) and the University of Cambridge (K. Bandara, F. Cirak) we are working on a joint paper summarizing our results. The paper will be submitted to an impacted journal and except for the methods described here it will include a detailed section devoted to the multiresolution analysis also used in the numerical examples given above. In addition to the above presented results the paper will also include a real-life problem that was solved in a cooperation with the R&D department of ABB in Baden-Dättwil, Switzerland. Talks related to this research have already been given by the author at

- TU Graz within an internal seminar on 17/1/2013 and 10/7/2013,
- the closing conference of the CASOPT project held in Baden-Dättwil, Switzerland, on 12/3/2013,
- the MAFELAP conference held in Brunel, London, UK, on 14/6/2013.

References

1. Delfour, M. C., Zolésio, J.-P.: “Shapes and Geometries: Metrics, Analysis, Differential Calculus, and Optimization”, SIAM, 2011, 621 p.
2. Greengard, L., Rokhlin, V.: “A Fast Algorithm for Particle Simulations”, *J. Comput. Phys.* 73, 1987, p. 325–348.
3. Eppler, K., Harbrecht, H.: “Tracking Neumann Data for Stationary Free Boundary Problems”, *SIAM J. Control Optim.* 48/5, 2009, p. 2901–2916.
4. Haslinger, J., Mäkinen, R. A. E.: “Introduction to Shape Optimization: Theory, Approximation, and Computation”, SIAM, 2003, 273 p.
5. Haslinger, J., Ito, K., Kozubek, T., Kunisch, K., Peichl, G.: “On the Shape Derivative for Problems of Bernoulli Type”, *Interfaces and Free Boundaries* 11/2, 2009, p. 317–330.
6. Hsiao, G. C., Wendland, W. L.: “Boundary Integral Equations”, Berlin: Springer-Verlag, 2008, 618 p.
7. Meyer, M., Desbrun, M., Schröder, P., Barr, A. H.: “Discrete Differential-Geometry Operators for Triangulated 2-Manifolds”, *Visualization and Mathematics III*, Springer, 2002, p. 113–134.
8. McLean, W.: “Strongly Elliptic Systems and Boundary Integral Equations”, Cambridge: Cambridge University Press, 2000, 357 p.
9. Of, G., Steinbach, O., Wendland, W. L.: “The Fast Multipole Method for the Symmetric Boundary Integral Formulation”, *IMA J. Numer. Anal.* 26/2, 2006, 272–296 p.
10. Sauter, S. A., Schwab, C.: “Boundary Element Methods”, Berlin: Springer-Verlag, 2011, 561 p.
11. Sokolowski, J., Zolésio, J.-P.: “Introduction to Shape Optimization: Shape Sensitivity Analysis”, Berlin: Springer-Verlag, 1992, 250 p.
12. Steinbach, O.: “Numerical Approximation Methods for Elliptic Boundary Value Problems: Finite and Boundary Elements”, New York: Springer-Verlag, 2007, 279 p.
13. Steinbach, O., Rjasanow, S.: “The Fast Solution of Boundary Integral Equations”, New York: Springer-Verlag, 2008, 386 p.
14. Zapletal, J.: “The Boundary Element Method for the Helmholtz Equation in 3D”, MSc. thesis, Department of Applied Mathematics, VŠB-TU, Ostrava, 2011, 90 p.

The convergence of OpenBTS and OpenIMSCore projects

Marcel Fajkus, Jakub Cubik, and Stanislav Kepak

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{marcel.fajkus, jakub.cubik, stanislav.kepak}@vsb.cz

Abstract. This article deals with the open source projects that implement software based GSM and IMS technologies. OpenBTS project is software that allows creating your own GSM network, which can be reached via GSM mobile phones. OpenIMSCore project implements the core of IMS network, which is based on IP network. An interesting solution is to integrate these two technologies allowing the GSM end users interconnection with IMS end users who use the Internet to make calls using VoIP technology.

1 Introduction

IMS Technology was established by 3GPP organization and is a major part of the packet networks that combine voice, data and multimedia networks. German Institute FOKUS has designed and implemented an open source solution of IMS core architecture called OpenIMSCore [1]. This solution is used for testing and scientific research purposes. OpenIMSCore allows you to call other users with IMS clients not only with each other in the local network, but also make calls to users outside the local network, if OpenIMSCore is connected to the Internet.

OpenBTS is an open source application that implements a digital cellular radio network GSM [2]. To create a full-GSM Um interface OpenBTS uses software-controlled universal radio USRP. Asterisk PBX provides call routing [3]. Users with GSM mobile phones can register to the private GSM network and call to each other. Interconnection of both projects can create an extensive network that is able to connect mobile users using standard GSM phones with users connected to the Internet using IMS account.

2 OpenIMSCore

OpenIMSCore is a project that implements the core of IMS architecture and is used for testing purposes of this technology.

In 1988, the Institute FOKUS in Berlin was founded [4] in order to develop communication systems. One of its many projects is OpenIMSCore. OpenIMSCore project consists of two parts. The first part consists of ser_ims [5], which implements CSCF (Call Session Control Function) servers. These servers provide a call control and signaling. Second part consists of FHoSS [6], which implements the HSS server (Home Subscriber Server).

P-CSCF, I-CSCF and S-CSCF elements have been developed as an extension of SER (SIP Proxy Server), because the SER is a high performance, configurable SIP server licensed under the GNU, which can act as a SIP registrar, proxy or redirect server. [5]. OpenIMSCore form the core of IMS architecture, which can be used for testing and scientific research purposes. Typical scheme of the test environment [6] is shown below.

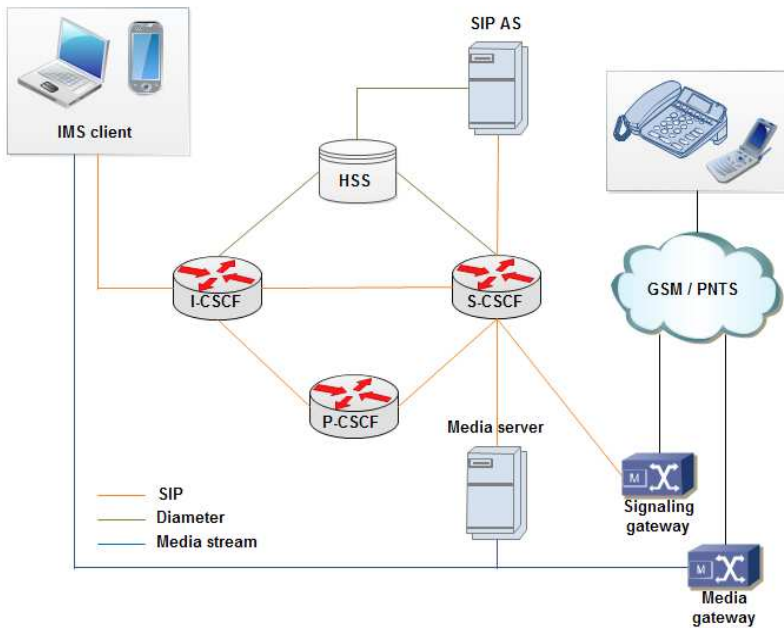


Fig. 1. Typical scheme of the IMS network

IMS client communicates with the outer world via I-CSCF server. I-CSCF addresses specific S-CSCF based on the record in the HSS, then provides registration and connects clients IP address together with its SIP address. When there is a request for setting up a call, the IMS client sends a request which is routed by P-CSCF elements to the destination I-CSCF element of the second user. If the call is routed to the GSM or PSTN network signaling and transport protocols are transformed.

Architecture itself uses the SIP protocol for sending signaling messages between end user’s IMS and components I-CSCF, S-CSCF and P-CSCF. Diameter protocol is used for providing AAA services that include authentication, authorization, and accounting. The actual media data transmission is directly transmitted between IMS clients and in case of calls to the PSTN or GSM network is transmitted through the media gateway.

3 OpenBTS

OpenBTS (Open Base Transceiver Station) is an open-source Unix application that allows you to create your own GSM network using the universal transmitter USRP from Ettus Research [7]. OpenBTS uses a universal software controlled radio USRP, it is connected to a host computer that controls the transmitting and receiving data in the USRP device. The USRP universal transmitter produces full air interface Um. Asterisk PBX software is used for call routing.

The project was initiated in order to reduce the cost of providing GSM services in rural areas, developing areas and difficult terrains such as oil platforms, etc. The public release of OpenBTS is the first freely distributable software that implements the lowest three layers of the industry-standard GSM. OpenBTS is written in C++ and released as free software under AGPLv3 license.

OpenBTS does not redirect speech traffic through the mobile network operator, but delivers this traffic using SIP protocol via VoIP PBX software. Um air interface is implemented by software defined radio transmitter (USRP) from Ettus Research, so there is no need for a dedicated GSM hardware.

Um interface is a wireless interface of GSM mobile phone standard. It is the interface between the mobile station (MS) and the base transceiver station (BTS). It is called Um, because it is a form of mobile ISDN U interface. Um also supports GPRS packet-oriented communication.

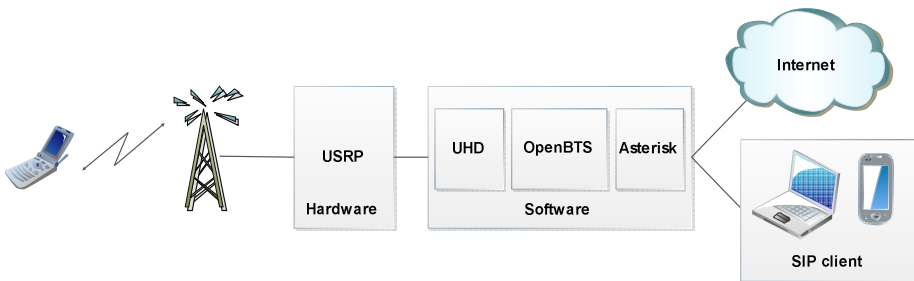


Fig. 2. The structure of GSM network created using OpenBTS and USRP

In OpenBTS entire GSM network hardware is replaced by USRP transmitter. USRP receives and transmits GSM signaling and is controlled by software driver UHD. OpenBTS fulfills the role of MSC / VLR and Asterisk represents PBX that provides call routing.

4 Convergence of OpenBTS and OpenIMScore projects

By combining of both projects network that enables convergence of mobile users with GSM phones and IMS users can be assembled. Scheme of such a network is in Figure 3.

Open-bts.home server runs project OpenIMSCore and DNS server to translate domain names. IMS client Monster with an account sip: 2000@open-ims.home is installed on this server. On OpenIMSCore are registered users, whose numbers meet the shape 2XXX.

Open-bts.home server is running Asterisk and OpenBTS, which uses sqlite3.db database. This database is used for mapping mobile phones IMSI numbers to SIP accounts. Asterisk then performs call routing using the SIP accounts. Server is simply connected to the USRP via Ethernet. Mobile users in OpenBTS have assigned SIP accounts 3XXX. The particular mobile phone with IMSI00101xxxxxxxxx has assigned SIP account sip: 3000@open-bts.home.

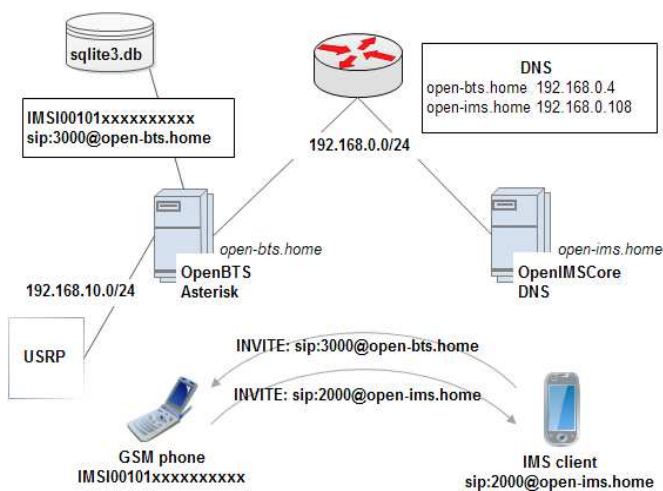


Fig. 3. Network scheme, which integrates GSM and IMS projects using the OpenBTS and OpenIMSCore

If the user on his GSM phone dials the number 2000, the request is sent using the GSM network via USRP to the OpenBTS server. Server then provides routing with Asterisk PBX. Asterisk based on its configuration forwards the INVITE to the open-bts.home server, which further forwards the INVITE to the IMS client with particular public address.

If the IMS user with the number 2000 wants to make a phone call with GSM user, he will dial 3000. Dialing the telephone number INVITE request will be sent to the home server I-CSCF. Based on configuration of OpenIMSCore the call is routed to the open-bts.home server. Asterisk running on open-bts.home gets the IMSI of GSM phone based on the dialed number 3000 (database sqlite3.db) and sends a request on call to the end GSM phone using the Um interface.

5 Conclusion

In this paper IMS network architecture using the OpenIMSCore project was made. IMS users reached with each other and also reached to SIP users in case of IMS

network connected to the Internet. Private GSM network was also implemented, the mobile users can make calls with each other and send SMS messages. The last step was the setting that led to the convergence of both technologies. As a result, it is possible to realize calls between multiple users across the IMS and GSM technology. In this article interconnection of both technologies has been described and implemented.

We will deal with achievable speech quality for actual calls, limits of simultaneous calls for individual projects, etc. At this point, we are already working on these measurements. The results will be presented at the conference KTTO in late 2013.

References

1. Welcome to Open IMS Core's Homepage. [online]. [cit. 2013-07-16]. Available from: <http://www.openimscore.org/>
2. GNU Radio: The OpenBTS Wiki Subspace. In: [online]. [cit. 2013-07-16]. Available from: <http://gnuradio.org/redmine/projects/gnuradio/wiki/OpenBTS>
3. GNU Radio: Setting Up Asterisk to Work with OpenBTS. In: [online]. [cit. 2013-07-16]. Available from: <http://gnuradio.org/redmine/projects/gnuradio/wiki/OpenBTSSettingUpAsterisk>
4. OSIMS - The FOKUS Open Source IMS Core. In: [online]. [cit. 2013-07-16]. Available from: http://www.fokus.fraunhofer.de/en/fokus_testbeds/open_ims_playground/components/osims/index.html
5. The Open IMS Core CSCFs. [online]. [cit. 2013-07-16]. Available from: http://www.openimscore.org/docs/ser_ims/index.html
6. The Open IMS Core HSS. [online]. [cit. 2013-07-16]. Available from: <http://www.openimscore.org/docs/FHoSS/index.html>
7. USRP Embedded Series. [online]. [cit. 2013-07-16]. Available from: https://www.ettus.com/product/category/USRP_Embedded_Series

Visible Light Communication with Four Wavelengths Division Access

Stanislav Hejduk, Jiří Bocheza, Karel Witas, and Artem Ganiyev

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{stanislav.hejduk, jiri.bocheza, karel.witas, artem.ganiyev}@vsb.cz

Abstract. In this paper we describe possibilities of visible light communication (VLC) system with simultaneous communication through 4 different wavelengths. We also consider different approaches to modulation schemes and its possible influence on multipath signal propagation. Our objective is to maintain constant white color of the light with presence of communication signal. To preserve the simplicity and versatility of the modulator we designed modulator which can be powered together with modulated light. This solution is advantageous for its simplicity and especially for its low price.

1 Introduction

LED diodes are significant part of our everyday life for many years. Thanks to falling prices exceptional lifetime and increasing optical power we can expect that LEDs will replace fluorescent and incandescent light bulbs in few years.

Although the idea of contemporary lighting and communication isn't new [1] this technology is still suppressed by radio communications for its higher data rates. However, future development of this technology is still very promising, especially for its coverage parameters. There is possibility to precisely define covered area and use more identical systems in one area even without dividing it by walls or obstacles [2], [3].

For our type of Visible Light Communication (VLC) system, will be covered area considered as diffusive network. This means that light is evenly flooding the covered area with high optical power and signal can be detected also after reflection from walls and other obstacles, so the communication is possible even behind direct Line-Of-Sight (LOS) conditions.

Of course we must not forget speed limitations in order to multipath signal propagation which causes expansion of transmitted pulses and for higher frequencies even their multiplication. As a result the photodetector will receive same data several times in succession. [4]

Nowadays are for this purpose used mainly white LEDs, especially for their relatively wide wavelength spectrum. This is caused by luminophor which is activated by

blue LED base on the chip. [5] This luminophor however reduces modulation speed in fluorescent part of the spectrum.

We want to circumvent this limitation by using combination of RGBY (Red-Green-Blue-Yellow) LEDs. This solution can obtain four-times more data in the same time period and thanks to the yellow colour added in the spectrum possibly create more convenient white light than with classic RGB combination.

2 RGBY model

Since the usual RGB model covers only small part of the light spectrum there are places for another wavelengths. In this case was chosen yellow (Y) color. It should help to create so called warm light and fill in one empty place in optical spectrum. Fig. 1 shows an example of this case.

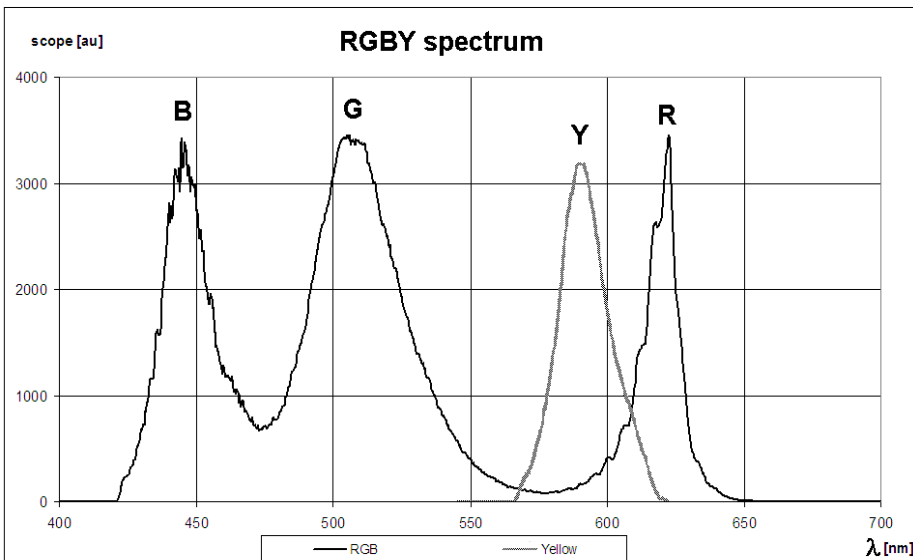


Fig. 1. RGBY signal spectrum.

Unfortunately, we can see that light spectrum is noticeably shifted to red part of the spectrum. As a result we have to precisely define light intensity for all four colors. This is not unexpected problem, but lower light intensity means also smaller communication distance or lower communication frequency.

Compared with daylight spectrum and light spectrum of the white LED with luminophor (shown in Fig.2) there is still empty window in wavelengths around 550 nm. In case of daylight there is also minor IR part of the spectrum missing. IR part of the spectrum is deliberately left free for uplink communication because background light can be also defined as noise which causes problems with SNR ratio. For mobile devices is also necessary to save more energy and therefore is necessary to communicate in less noisy environment.

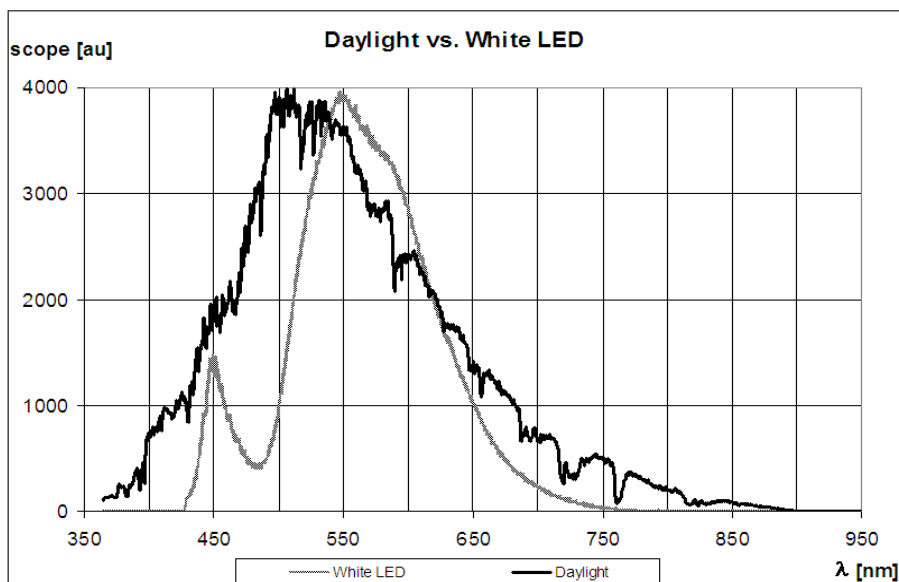


Fig. 2. Comparison of Daylight and White LED spectrum.

2.1 Construction

Mechanical construction is based on 4 high power LEDs placed close together on common heat sink (as shown in Fig.3).

Problem is that colors have to be precisely mixed at every part of the covered area. Therefore network should be diffusive to ensure appropriate area coverage. This brings the need for a suitable diffuser which will be able to mix the light before it leave the light source and also remove potential high intensity spots. Otherwise we will experience multicolored shadows on the walls and other obstacles.

Another part of the problem is the loss of light intensity caused by diffuser. Since we have to keep output power in applicable values we consider use of soda lime ground glass diffuser which should have better scatter loss parameters than common plastic materials.

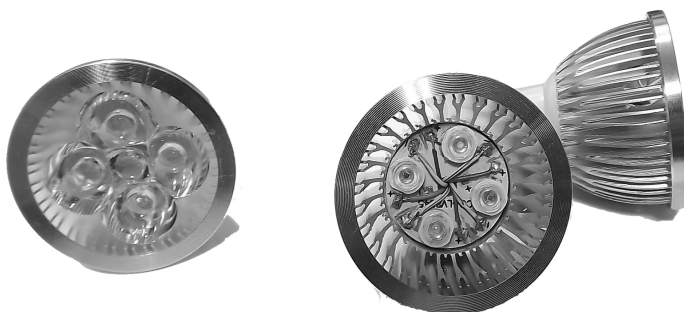


Fig. 3. RGBY LED construction with collimating optics (left) and without (right).

2.2 Modulation

Basic requirement is to keep a stable mean value of the transmitted signal. For simple On-Off-Keying modulation (OOK) we just turn light ON for logical 1 and turn light OFF for logical 0. It means that the logical levels at the input are simply transferred to the output. So the long sequence of zeros causes loss of light, which is bad especially in VLC systems. Therefore with OOK modulation is necessary to use specific coding scheme in order to keep stable mean value of the signal.

All schemes mentioned in Fig.4 keep mean value stable for all possible data sequences and makes communication unobservable by human eye with simple OOK modulation.

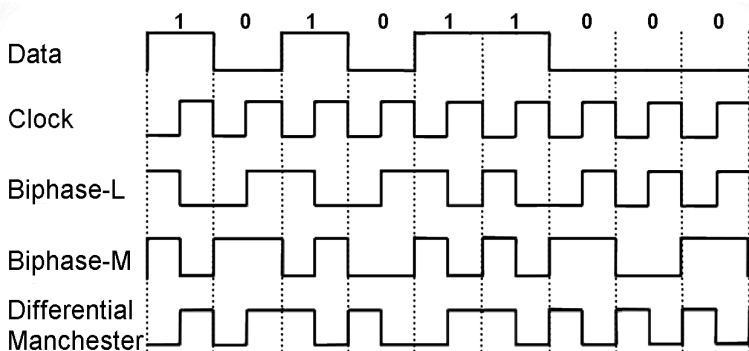


Fig. 4. Coding schemes suitable for VLC with OOK modulation.

Now we face the question of using mentioned 4 channels.

First idea is to use each channel independently and have four individually communicating devices. Advantage of this solution is easier simultaneous communication with elimination of collisions caused by multiple accesses to common channel. However, communication speed will be smaller.

Second idea is parallel sending of data with all LEDs simultaneously. It means that we can transmit 4 individual bits in one clock period. Another advantage is possible multipath signal propagation elimination because we can change communication wavelength with every single bit and therefore significantly reduce errors caused by reflections from distant obstacles. As a result we don't need to use coding methods like Dynamic OCDMA [6] that eliminates multipath signal propagation by sending every block of bits with individual code sequence.

3 Modulator and measurement

Construction of the modulator should accomplish number of requirements. It should be simple and easily integrable to the light source because of possibility to use existing LED lights and power sources. It should be as fast as possible and of course LOW-COST construction is preferred. However, these requirements are mutually exclusive and it is therefore necessary to choose a suitable compromise.

For initial testing of different LED light sources was build modulated current source which allows supply different number of LEDs connected together for higher optical power output. Modulator scheme is shown in Fig.5.

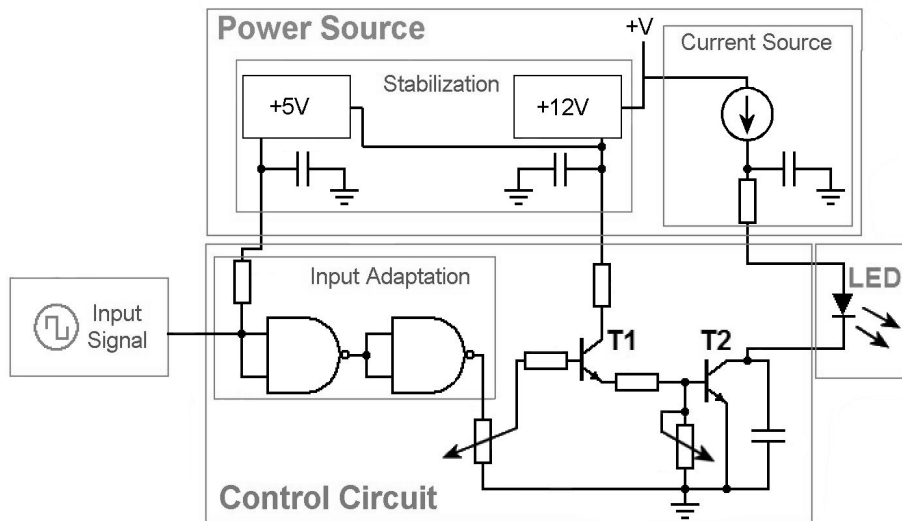


Fig. 5. Block scheme of current modulator.

The main advantage is that this modulator is able to operate with only one simple power source. Stabilization is, together with adjustable current source, placed in the power source part. Control circuit is simple and consists of two transistors with adjustable operating point and input adaptation circuit.

Nevertheless, the price for this simplicity is lower performance and working frequencies are only around 3MHz. In following figure (Fig.6) are values of light output intensity for different modulation frequencies and different operating point settings.

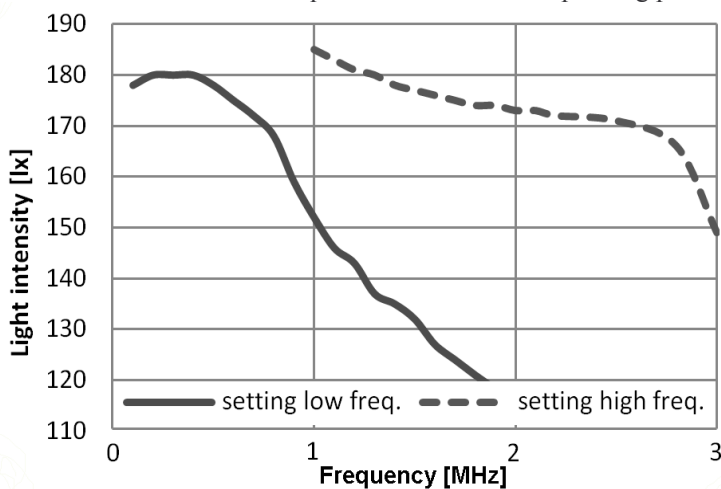


Fig. 6. Dependence of output light intensity and input frequency.

4 Conclusion

Although 3MHz limit may seem to be small, if we take into account background noise and limits of the photo detectors sensitivity caused by higher load values, reached frequency is for mobile diffusive network quite sufficient.

If we forget limits of the used modulator, remaining high power LEDs have also higher capacity of PN junction which causes slower reaction times. To avoid this consequence we can use larger number of smaller LEDs instead of one high power LED or we can use LASER, but this solution could be unsafe for human eyes.

In fact the most problematic part is to ensure uniform diffusion of light. For this multi-wavelength system is collimating optics usually inapplicable because of multi-color high intensity spots.

References

1. T.D.C.LITTLE, P.DIB, K.SHAH, N.BARRAFORD, B. GALLAGHER. Using LED Lighting for Ubiquitous Indoor Wireless Networking. IEEE International Conference on Wireless & Mobile Computing, Networking & Communication, 2008.
2. Jayasri AKELLA, Chang LIU, David PARTYKA, Murat YUKSEL, Shivkumar KALYANARAMAN, and Partha DUTTA. Building Blocks for Mobile Free-Space-Optical Networks. Proceedings of IFIP/IEEE WOCN, Dubai, United Arab Emirates, March 2005.
3. Dayong ZHOU, Peter G. LoPresti, Hazem H. REFAI. Enlargement of Beam Coverage in FSO Mobile Network. Journal of Lightwave Technology, vol. 29, no.10, MAY 15, 2011.
4. Adrian MIHAESCU, Pascal BESNARD. Indoor Wireless Optical Communications. 8th International Conference on Communications, Bucharest, Romania, 2010.
5. Jelena GURBOR, Sian Chong Jeffrey LEE, Klaus-Dieter LANGER, Ton KOONEN, Joachim W. WALEWSKI. Wireless High-Speed Data Transmission with Phosphorescent White-Light LEDs. Optical Communication – Post-Deadline Papers (published 2008), 2007 33rd European Conference and Exhibition of, Berlin, Germany, Sept. 2007.
6. R. LUCACIU, A. MIHAESCU, C. VLADANU, Dynamic OCDMA coding for indoor wireless optical communications, Communications (COMM), 2010 8th International Conference on , 10-12 June 2010.

Acknowledgements

This article was created with the active support by the Ministry of Education of the Czech Republic within the projects no. SP2013/88 and SP2013/69 of the VSB-Technical University of Ostrava.

Measurement of homogeneity of the microwave emitter

Jakub Jaros, Andrej Liner, and Martin Papes

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jakub.jaros, andrej.liner, martin.papes}@vsb.cz

Abstract. The paper deals with an experimental verification of the homogeneity of the electromagnetic field of the microwave band (2.45 GHz). The verification of homogeneity consists in experimental measuring of the temperature distribution within the wooden sample. The method is based on heat exchange in an isolated system where wooden sample serves as an absorber of the irradiated energy. To identify locations with different power density was used DTS system, based on nonlinear phenomena in optical fibers.

1 Introduction

Many teams of researchers have been dealing with the problem of homogeneity of EM fields since the beginning of the use of microwaves. With this problem were dealing researchers like for example Watanabe ET al [1] and Watanabe and Suzuki, who analytically calculated the distribution of the power density of the electric field (EF) within the applicator. Kayshyap and Wyslouzil [2] successfully used to increase the homogeneity of the EM field special design of the waveguide system. With the issue of distribution of EM fields in the workspace of MW devices deals also extensive review by Li ET al [3]. Li ET al, among other things also summarize the two basic approaches how to eliminate this negative effect. These methods were also not very effective. Measurements took a very long time and likewise the computations were very difficult. Another disadvantage of these methods was the impossibility of the evaluation of the measured results in real/time.

Nowadays, new materials and technologies allow completely different approach in addressing these issues. One of the few technologies, which are resistant to electromagnetic radiation is the use of optical fibers. Optical fibers are mostly used for telecommunication purposes, but likewise it is largely used in sensors. Their main advantage is low price, small size and the electromagnetic resistance. These are the reasons, why our team decides to use, in the cooperation with Thermo Sanace s.r.o.

Company, provide the measurement of the homogeneity of the microwave emitter by using optical fibers and Raman DTS device.

2 Theory about Microwave Heating

The distribution of the EM field is described by Maxwell equations. Maxwell's equations are considering the foundation of electrodynamics and describe the EM field in each point of the space. The differential form of Maxwell's equations for constant permittivity and permeability without the use of the source is described as follows [4]:

$$\nabla \vec{E} = 0, \quad (1)$$

$$\nabla \times \vec{E} = -j\omega\mu\vec{H}, \quad (2)$$

$$\nabla \vec{H} = 0, \quad (3)$$

$$\nabla \times \vec{H} = \sigma\vec{E} + j\omega\epsilon\vec{E}, \quad (4)$$

where E is the electric field vector ($V\ m^{-1}$), H is the magnetic field vector ($Wb\ m^{-1}$), ω is the angular frequency ($rad\ s^{-1}$), μ is magnetic permeability ($H\ m^{-1}$), ϵ is the dielectric permittivity ($F\ m^{-1}$), $j = \sqrt{-1}$, and σ is electric conductivity ($\Omega^{-1}\ m^{-1}$).

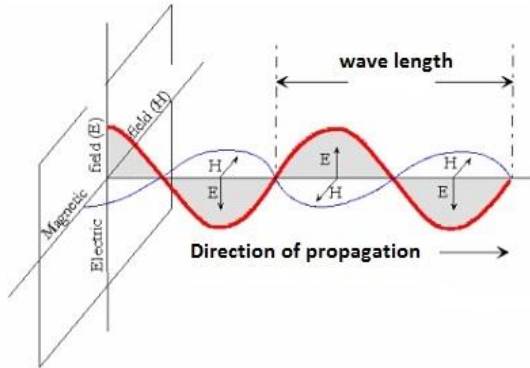


Fig.1. Electric and magnetic field with the direction of the distribution [7].

Vectors (\vec{E} a \vec{H}) are perpendicular to each other, have a consistent phase and ongoing oscillations perpendicular to the direction in which the wave propagates, as shown in Fig. 1. The direction of the wave's propagation is given by the vector multiplying $\vec{E} \times \vec{H}$. EM radiation is transverse radiation, which has wave properties (reflection, refraction, interference, polarization) and quantum / the photoelectric effect.

In the case that the wave propagates in the z direction, it can be described by the equation of the plane harmonic wave [5]:

$$E(x, c, z, t) = E_0(x, y, z)e^{j\omega t}, \quad (5)$$

$$H(x, c, z, t) = H_0(x, y, z)e^{j\omega t}. \quad (6)$$

3 DTS with Optical Fiber

The Optical Fiber DTS (Distributed Temperature System) are unique distributed temperature systems using optical fiber as a sensor. Temperature values are recorded along the optical fiber continuously in points. DTS system can be imagined as several thousand sensors providing information on the thermal state of the environment in which the optical fiber is located. These systems are mainly due to its advantages, utilized in many applications. The biggest advantages are:

- Resistance to electromagnetic radiation.
- Resistant to aggressive environments.
- The length of the measured section up to the 30 km.

As the name suggests, Optical Fiber DTS based on Stimulated Raman Scattering are using nonlinear Raman scattering. Lasers used in these systems operate at a wavelength of the 1064 nm. Raman spectra peaks are in this case shifted by ≈ 40 nm. That is equal to 1104 nm and 1024 nm. These two newly incurred components that arise from the reflections on the core and cladding boundary along the optical fiber are two parts of the spectrum and named as Stokes and Anti-Stokes component. Exactly the Anti-Stokes spectra component changes its intensity depending on the temperature along the fiber. The Stokes part of the spectrum is thermally independent. The DTS defines the location of temperature based on changes in the intensity of the Anti-Stokes spectrum and final ratio between Stokes spectrums [6].

Spatial resolution of the DTS system is standardly about 1 m with accuracy of $\pm 1^\circ\text{C}$, at a resolution of $0,01^\circ\text{C}$.

4 Measurement and Result

Measurements were performed on the experimental workplace at VSB-TU Ostrava. For measurements was used the microwave emitter with a magnetron type NL 10250. This emitter operates at the frequency of 2,45GHz and at the wavelength of the EM wave 12,25 cm. The listed maximum power is 1,9kW. Antenna for directing the flow of energy has a funnel shape with an aperture of 270 x 270 mm, as shown in Fig. 2.

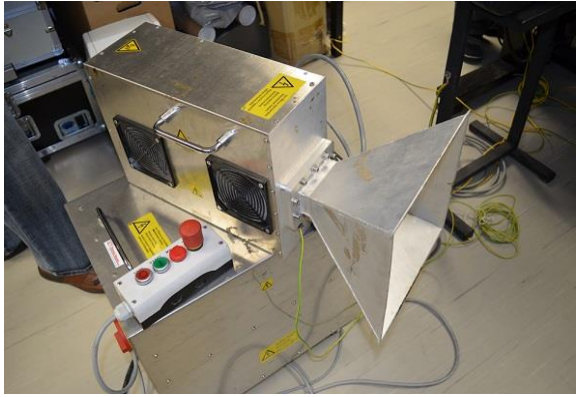


Fig.2. Microwave emitter with magnetron type NL 10250.

For measurements was made a board with dimensions 400 x 400 mm. On this board was created a matrix of optical sensory rings. The dimensions of this matrix were 270 x 270 mm and corresponded to the exact dimensions of the emitter’s antenna. Matrix contained 36 optical sensory rings. Each one of these rings had a small diameter of 30 mm and contained 3 m of optical fiber. The diameter of the ring and the length of the optical fiber in it were determined from the results of previous experimental measurements. Matrix shown in Fig. 3.

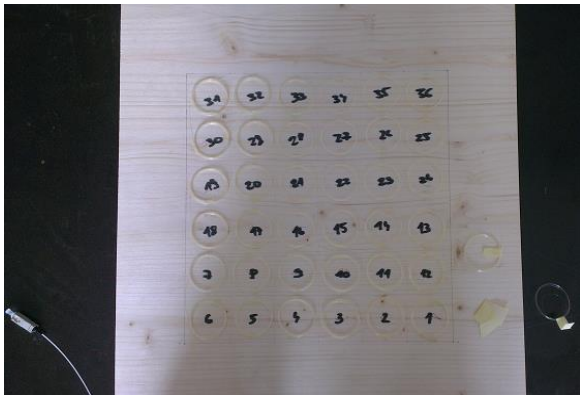


Fig.3. Measurement sample.

The measurement was carrying out in the time interval of 10 minutes. The measuring device DTS was set to 10 seconds iterations. The initial temperature of the sample was 23°C and during its heating, some parts of this sample were heated to the temperature about 86°C. In Fig. 4 can be seen, that the heating by the microwave emitter with magnetron type 10250 was inhomogeneous. Many parts were radiated by varying sensitivity. In addition, it can also be seen that the edges of the sample were heated not at all. It can be said, that the emitter is not radiating at the entire area of it’s antenna, but only at the area of 270 x 180 mm. It can be also seen, that the biggest heating was in the middle of the sample. This can be seen in individual parts Figure 4.

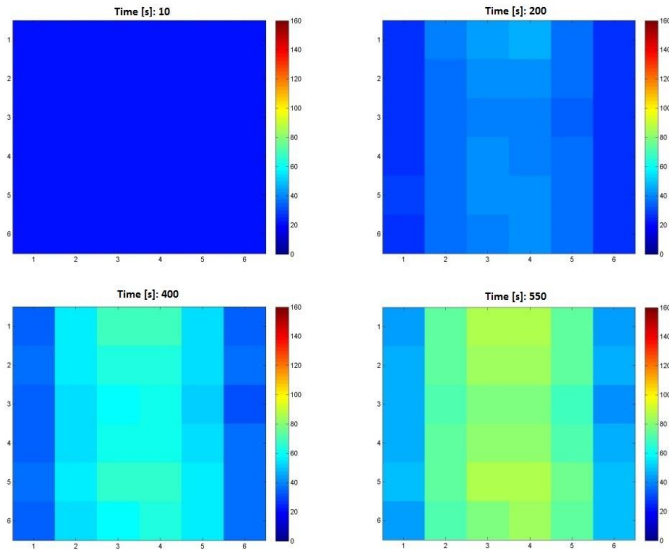


Fig.4. Measurement sample at a different time.

5 Conclusion

The aim of this measurement was the experimental verification the homogeneity of the radiated field microwave emitter and the verification of the functionality of the method using optical fibers and Raman DTS device in this and also similar applications. The results of the measured EM filed are presented above. In the graphs, it can be seen, that the EM field of the microwave emitter is really inhomogeneous.

References

1. Watanabe, M.; Suzuki, M.; Sugimoto, K. Theoretical and experimental study on uneven heating in microwave oven. In proceedings of the 1971 Symposium on Microwave Power, Vol. 1, Edmonton, Canada, May 26–28, 1971, 26–28. Baldonado, M., Chang, C.-C.K., Gravano, L., Paepcke, A.: The Stanford Digital Library Metadata Architecture. *Int. J. Digit. Libr.* 1 (1997) 108–121.
2. Kashyap, S.C.; Wyslouzil, W. Methods for improving heating uniformity of microwave ovens. *Journal of Microwave Power and Electromagnetic Energy* 1977, 12(3), 223–230.
3. Risman, P.O.; Ohlsson, T.; Wass, B. Principles and models of power density distribution in microwave oven loads. *Journal of Microwave Power and Electromagnetic Energy* 1987, 13(4), 193–198.
4. Kanarev, M., 2004: The Foundations of Physchemistry of Microworld. The fifth addition. Krasnodar. 400 pages.

5. Brodie, G., 2007: Simultaneous Heat and Moisture diffusion during microwave heating of mosit wood. American Society of Agricultural and Biological Engineers Vol. 23(2): 179-187. ISSN 0883-8542.
6. Koudelka P., B. Petrujova, J. Latal, F. Hanacek, P. Siska, J. Skapa and V. Vasinek. Optical fiber distributed sensing system applied in cement concrete commixture research. Radioengineering. 2010, vol. 19, no. 1, pp. 172-177. ISSN 1210-2512.
7. Sadiku, M.N.O., 2001: Elements of Engineering Electromagnetics. In: Brodie, G. (Ed.). Simultaneous Heat and Moisture diffusion during microwave heating of mosit wood. American Society of Agricultural and Biological Engineers Vol. 23(2): 180. ISSN 0883-8542.

Integration of voice services into corporate information systems

Martin Mikulec, Jan Rozhon, Filip Řezáč,
Jiri Vychodil, and Miroslav Vozňák

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{martin.mikulec, jan.rozhon, filip.rezac,
jiri.vychodil, miroslav.voznak}@vsb.cz

Abstract. The paper deals with an integration of voice services granted by Private Branch Exchange Asterisk into other information and communication systems. Results of this interconnection are advanced services. These advanced services improve speed and quality of exchanging information between users. This advanced services are taught in Voice over IP lessons in our department. Student's task during a lesson is to configure advanced services on assigned virtual servers according to prepared manual.

Keywords: Asterisk, XMPP, Calendar, Advanced services, Openfire IM.

1 Introduction

Modern trend in telecommunications today is an integration of voice services into other information and communication systems. The typical example of this integration is Elastix [1] project. It combines Asterisk PBX with Instant Messaging Server, Mail Server and many other services into one server.

The Asterisk PBX was chosen because of its open source framework and huge developer community. It guarantees a large amount of extensions and plugins for interconnection to other applications.[5] Asterisk is widely spread. It powers IP PBX systems, VoIP gateways, conference servers etc. [4]

XMPP (Jabber) is an XML based protocol used primarily for presence and messaging. It is an open standard, and there are several open server implementations such as Ejabberd or Openfire as well as several open source clients. Both server and clients are under Jive Software company development, so there is a good compatibility between the clients and the server. We have used an Openfire server project [2] with Spark and SparkWeb clients.

Asterisk was connected with calendar server too. We have chosen Davical calendar server. Davical implements the CalDAV protocol, which was designed for saving the contents of a calendar in the iCalendar format. The content in this format is saved on a remote shared server. Users can access these calendars with calendar clients. Currently, CalDAV is a standard RFC 4791, and therefore it supports most of

the standard calendar clients such as Mozilla Calendar (Sunbird / Lightning), Evolution, Mulberry, Chandler or iCal. This open source application uses a SQL database and friendly graphical interface, which defines the read or write permissions. There was Sunbird client used to access the calendar. Both local and network calendars are supported by this client.

We have connected all these servers together, so new interesting services were created.

Firstly, we have installed Asterisk on a virtual server with two clients connected to it. Then the calendar and jabber servers were installed and configured to work properly together. All new services are presented below.

The main advantage of this solution, in comparison with commercial solutions presented by Microsoft Lync or Siemens Unified Communication server, is the fact, that the whole solution is based on opensource services. The source codes of all applications are open and there is no problem with a modification of any part of the system.

Students during a lesson have a chance to install, configure and debug whole system from beginning to complete working system. They can see and understand main principles and protocols, which is the system based on.

2 Calendars

- Checking for user status (etc: available, busy, not disturb)
- Calendar entry (etc: Call Detail Record)
- Listen calendar content using module for Text-To-Speech

2.1 Installation of Davical server

Installation of Davical server consists of a few steps: [6]

- Installation of davical package and postgresql database server with corresponding modules
- Configuration of Davical database
- Creating of new apache virtual host for running web administration of davical server
- Creating of users and relevant calendars on davical server

Then, the created calendars are available on server under http://IP_address_of_the_server/caldav.php/user/home/ http link. These data can be used in any calendar client that supports CalDAV protocol. In our case we have used open source calendar client Sunbird.

2.2 Asterisk and Davical cooperation

Asterisk can read and write to the established calendar, but before it is necessary to modify file `calendar.conf` and connect asterisk with calendar.

2.3 Applications of calendar function

`CALENDAR_BUSY()` function returns the currently set availability, which the user has set in the calendar. This condition is controlled by the set interval in `calendar.conf`. The extension 666 was added at the end the file.

```

exten =>
666,1,GotoIf(${CALENDAR_BUSY(asterisk)}?busy:free)
exten => 666,n(busy),Dial(SIP/102)
exten => 666,n,Hangup()
exten => 666,n(free),Dial(SIP/101)
exten => 666,n,Hangup()

```

The call to extension 666 will have the following scenario. Status in calendar makes decisions. If available dialplan would be directed to 101, if status is busy, then at 102.

`CALENDAR_WRITE()` function is used to write to the calendar. This function is used to capture log entries. Call 777 cause a calendar entry, which includes information about when and who called. This log can be checked in the event listings for CalDAV server by Sunbird.

```

exten => 777,1,Set(start=${EPOCH})
exten => 777,n,Dial(SIP/101)
exten => h,1,Set(end=${EPOCH})
exten =>
h,n,Set(CALENDAR_WRITE(asterisk,summary,start,end)=Volal
$
${CALLERID(all)},{start},{end})

```

`CALENDAR_QUERY()` is a function that has two parameters the name of calendar and time in UNIX format. `CQ` returns the ID and event field, which will be used in function `CALENDAR_QUERY_RESULT`. This function accepts the following parameters: ID from `CALENDAR_QUERY`, requested information, the order of events.

The required information are summary, description, organizer, location, calendar, uid, start, end, busystate and attendees.

Any of the Text-To-Speech modules, such as Festival, is required to play the information.

CE is similar to CALENDAR_QUERY. This feature serves as a voice reminder of specific time before the event.

```
[calendar_event_notify]
exten => s,1,Answer
exten => s,n,Festival(You have an event in 10 minutes\,
which will be at $
{CALENDAR_EVENT(location)}\, object of event is $
{CALENDAR_EVENT(description)}\, bye)
    exten => s,n,Hangup
```

3 Jabber

We have chosen Openfire project as jabber server. Instalation was made in a few steps in web environment.

Than 2 client accounts (alice and bob) were hen created via XMPP clients Spark and Pidgin. After successful registration of clients, Asterisk was interconnected with Openfire throught Asterisk Management Interface (AMI). There is Asterisk-IM plugin in the Openfire. We have created a manager account in Asterisk and connected Openfire via this account.

Then we have mapped XMPP accounts to Asterisk extension.

After successful connection of clients, two new services were available.

3.1 Automatic change of XMPP status according to user's active call

Openfire automatically reacts to the user's state in Asterisk. The status on XMPP client is automatically changed to "On Phone" when user have incoming or outgoing call in Asterisk. This new feature can inform user's colleges in friend list that user has an active call but he is not fully prepared to communicate with them throught XMPP client.

3.2 Direct calling from XMPP client Spark

The second feature created by interconnection of Asterisk and XMPP server is a direct calling XMPP users from XMPP client Spark.

If a XMPP user has a mapping SIP account in Asterisk, he can be called directly from Spark.

1. Caller Alice finds called Bob in the friend list of her XMPP client Spark

2. She chooses “Call” option by Bob’s account
3. Asterisk will call Alice’s phone
4. When Alice receives the call, Bob’s phone will start ringing
5. When Bob receives a call, Alice and Bob are connected

The main advantage of this feature is, that there is no need to Alice to know Bob’s phone number. This number is stored in Openfire database and it is available for all Bob’s friends in a list.

4 Jabber functions

Asterisk can be connected to Openfire server as common client. Client has to be configured in jabber.conf file in Asterisk directory. The following parameters have to be specified there:

IP address of XMPP server

Connection port

Common status of Asterisk XMPP client (online)

Buddy list – it defines friend’s accounts that will Asterisk communicate with.

A typical application example is a notification for user about incoming call in XMPP client’s window [3]. The user can react and send the call to the proper destination.

1. Bob calls Alice, who is available under 333 extension number
2. Asterisk informs Alice in her XMPP client window about incoming call and offers her destination extension number, which she can send the call at.
3. Alice chooses any destination extension number in her XMPP client window.
4. Asterisk reads the Alice’s answer and sends the call to the proper extension number.

Example of implementation in dialplan

```

exten => 333,1,Answer();
exten => 333,n,
JabberSend(asterisk,alice@192.168.209.130,Incoming call
from ${CALLERID(name)}, choose destination);
exten => 333,n,
JabberSend(asterisk,alice@192.168.209.130,1 :Send to
101);
exten => 333,n,
JabberSend(asterisk,alice@192.168.209.130,2 :Send to
102);
exten => 333,n, Set(OPTION=${
{JABBER_RECEIVE(asterisk,alice@192.168.209.130)}});
exten => 333,n, GotoIf(${${OPTION} = 1}?10:20)

```

```

exten => 333,10,
exten => 333,n, Dial(SIP/101);
exten => 333,20,
exten => 333,n, Dial(SIP/102);

```

Next interesting function is routing the call according to a status of XMPP client. The standard defines 6 states of XMPP clients. They are:

Online, Free to chat, Away, Not Available, Do Not Disturb, Offline

Asterisk is able to route a call according to all these states.

1. Bob calls alice's extension number 555
2. Asterisk checks alice's status in her XMPP client
3. If alice is connected to Openfire (status 1-5), Asterisk sends the call to extension number 101
4. If alice is not connected to Openfire (status 6), Asterisk sends the call to extension number 102

Example of implementation in dialplan

```

exten => 555,1,Set(STATUS=${
{JABBER_STATUS(asterisk,alice@192.168.209.130)});
exten => 555,2,gotoif(${[${STATUS}] < 6}?3:10)
exten => 555,3,Dial(SIP/101)
exten => 555,10,(unavailable),Dial(SIP/102)

```

5 Conclusion

The new advanced services created by Asterisk-Openfire integration can be used in many scenarios. A typical application is in a call centre. There are many operators available under one shared public number and calls are routed among them according to many rules. One of the rule can be a status set in a XMPP client. The status can be set manually by the operator or automatically by PBX during an active call. The busy operator is not then able to answer calls, so the calls are routed to other available operators. An operator does not need to remember any phone numbers. It is able to establish the call directly from XMPP client. The phone numbers are stored in the XMPP server with users in a buddy list. The assistant is able either to answer the call or send the call to another operator by choosing the option in XMPP client. The last rule relates with incoming call handling. When a customer calls the call centre, the call can be routed according to the state in the calendar.

Acknowledgement

This material has been supported by FRVS grants No FRV 1467/2013/G1

References

- [1] E. Landivar, Unified Communications with Elastix vol.1 ,” in Creative Common Attribution, 2nd ed. vol. 1, 2011,
- [2] <http://www.igniterealtime.org/index.jsp>, Openfire and Spark documentation.
- [3] Vychodil, J., Mikulec, M., Safarik, J., & Voznak, M. (2012). Improvement of network efficiency on the grounds of change communication protocol. *Advances in Electrical and Electronic Engineering*, 10(4), 282-285
- [4] Vychodil, J., Tomala, K., & Voznak, M. (2010). Utilization asterisk for service desk. *Advances Electrical and Electronic Engeneering*, 8(3), 54-57.
- [5] GOMILLION, D.; DEMPSTER, B. Building Telephony Systems with Asterisk : An easy introduction to using and configuring Asterisk to build feature-rich telephony systems for small and medium businesses. Vyd. 2. Birmingham : Packt Publishing, 2005. 180 s. ISBN 1-904811-15-9.
- [6] Wiki.davical.org [online]. 2011 [cit. 2011-03-24]. Ubuntu/Lucid. Dostupné z WWW: <<http://wiki.davical.org/w/Ubuntu/Lucid>>.

Recognizing emotions from human speech using 2-D neural classifier and influence the selection of input parameters on it's accuracy

Pavol Partila, Karel Tomala, and Jakub Safarik

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{pavol.partila, karel.tomala, jakub.safarik}@vsb.cz

Abstract. In this article we give the comparison of different methods of speech features extraction for neural network classifier. Recognizing human emotion from speech is an important issue today, because of the massive deployment interface for communication with a computer using human speech. Therefore, automatic detection of emotional state of the speaker moved a huge step forward in the last few years and mainly in effort to increase accuracy, speeds and efficiency. For output-stage classifier was used self-organizing feature map, which is specific type of artificial neural nets. The number of input parameters must be limited for hardware and time consuming computation of neurons positions. Therefore we discuss about the accuracy of the classifier whose input is the fundamental frequency calculated by different methods. The result of this research deals with the accuracy of emotion classifier and compares the two input combinations.

1 Introduction

Secondary speech information is an important part of verbal communication. The way we pronounce a word carries a large amount of information. A different intonation of a spoken word changes the meaning. There are many words that mean something totally different with a different intonation. The word "immediately?" pronounced with a rising intonation at the end of the word has a different meaning than "immediately!" with the intonation at the beginning of the word. The intonation is a word associated with the emotional state of the speaker. There are many areas in which the information about the emotional state is needed.

Nowadays, technological development puts more emphasis on the increased accuracy and simplicity of communication between man and computer. Modern applications use the speech for input-output interface increasingly. In this type of interaction two problems can occur, caused by the absence of information about the emotional state. The first one is an incorrect recognition of a word or a command from a person who is under stress. The machine recognizes human speech differently than a man with his hearing. The accuracy is affected by changes in the voice signal due to stress on the vocal tract. The second problem is that we feel the absence of emotional state

in the machine speech of the loudspeaker. Classic applications such as Text-To-Speech combine parts of speech sounds that are truly correct, but ultimately this signal is without any emotion. Such speech acts on the man and is synthetically unreliable.

There are several physiological criteria such as for example heart rate, breathing changes and sweating, which enable determining the emotional state of a man. A number of speech signal parameters is used in speech processing. An imperfect human ear responds to parameters such as intensity, intonation, and speech rate. Fundamental frequency of speech, zero crossings rate, energy and cepstral coefficients are parameters which are used in digital speech processing. The following sections contain a description of signal processing methods, selection of the training sets and neural network classifier [1], [2], [6].

2 Fundamental frequency of vocal cords

Age, gender, speech errors and emotional state of a man can be determined using this parameter. There are several methods in signal processing which enable estimating the fundamental frequency. The vocal cords are almost completely open in the creation of voice-less phonemes. The basic tone does not arise with opened vocal cords and therefore F0 can be calculated only from the voiced parts. Each method to calculate F0 has advantages and disadvantages. This article compares four methods and their use in a self-organizing feature map classifier.

Auto-correlation function: This simple method can be used where we have a full signal. The auto-correlation function is defined in Equation 1.

$$R(m) = \sum_{n=m}^{N-1} s(n)s(n-m). \tag{1}$$

Lag k is determined by the position of maxima.

$$k = \arg \text{MAX}_m R(m). \tag{2}$$

The autocorrelation process is reversible and therefore the signal has a double length after this process. The distance between the first and the second peak is k. [4]. The disadvantage of this method is that the very high energy of original framework can absorb all the information from the other. The fundamental frequency of vocal cords is calculated using Equation 3.

$$F_0 = \frac{F_{\text{sampling}}}{k}. \tag{3}$$

Normalized cross-correlation function: The difference in energy shifted and original framework is settled by normalization:

$$NCCF(m) = \frac{\sum_{n=m}^{N-1} s(n)s(n-m)}{\sqrt{E_1 E_2}}. \tag{4}$$

Where E1 is energy of original frame and E2 is energy of shifted frame. Lag and F0 are calculated. F0 extraction is performed just as in the ACF. [3].

Auto-correlation with central clipping: Voiced and voiceless parts of the signal are separated with the threshold level. The speech signal constantly changes and thus it is not possible to use a single threshold for the entire signal. The threshold is determined for each frame separately, using the equation below.

$$P(i) = \alpha \text{MIN}(\text{MAX}_{i-1}, \text{MAX}_{i+1}). \quad (5)$$

Peaks of neighboring frames $i-1$ and $i+1$ are established. The threshold level is determined as the lower value of the two. Constant α is set within the range 0.8 to 1. The original signal of frame i is weighted with threshold $P(i)$. Samples with a higher value are clipped and normalized to 1 and lower to 0, as can be seen in Figure 1 below.

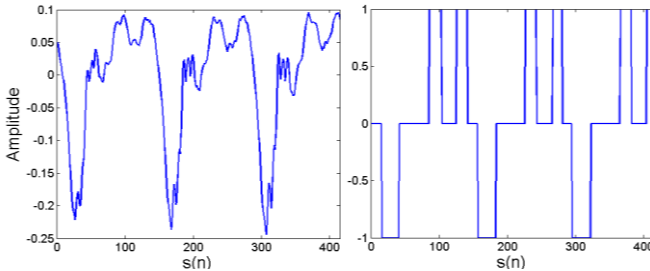


Fig. 1. Frame of speech signal before (left) and after clipping and normalization (right).

Next step is the auto-correlation and establishing the position of the second main lobe. F_0 calculation is simple and the same as described in previous methods. It is only necessary to find the position lag and use Equation 12 [4].

Subharmonic-To-Harmonic Ratio: This method is developed in frequency domain. $A(f)$ is short-term spectrum function. Suppose that f_0 is fundamental frequency. The sum of harmonic amplitude is defined below.

$$SH = \sum_{n=1}^N A(nf_0). \quad (6)$$

N is number of all harmonics which can be considered. If we consider the subharmonic frequency, that is the one half of f_0 .

3 Self organizing feature map

The emotional state classifier is based on self-organizing maps (SOM). These maps represent a specific type of neural networks with uncontrolled competitive learning. The learning process of SOM is uncontrolled which means that the input data do not need to know the output. In the process of learning SOM themselves determine how to classify the inputs. At the beginning of learning, the weights of all the inputs of all neurons can be set randomly. Randomly selected input vectors are applied to neurons and then analyzed in order to find one which is most similar to input. This neuron is called the winner. The weights of neighboring neurons are adjusted according to the following rule.

$$w_{ij}(t+1) = w_{ij}(t) + \gamma(x_j(t) - w_{ij}(t)). \tag{7}$$

This equation describes the weight between neurons i and j for $t+1$ iteration and input $x_j(t)$. Next iteration means a new vector on input, finding new winner and changing weights between neurons again. After the learning process is completed, the map has a shape that represents the characters of input parameters. The aim of this article is to clarify the appropriate choice of parameters for the two-dimensional maps. They are two-dimensional because counting the change weights between neurons for more than 2D dimensions poses high demands on hardware. This should be taken into account as these days real time speech processing is sought after in the entire telecommunications services sector [5].

4 Results

The system which was designed in software Matlab finally consists of three main blocks. These have been described above. Figure 2 shows the block diagram.

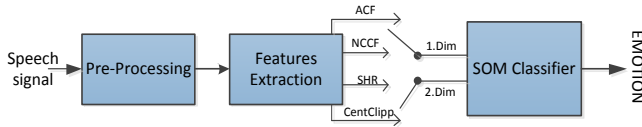


Fig. 2. Speech emotion recognition system

The input speech signal is embedded into the system sequentially. Audio records of different emotional states were used as the speech signals. The first block algorithm performed the pre-processing operations. The next block is designed to extract the basic tone using the four above-mentioned methods. These parameters are input data for the neural network. The combination of the four methods ACF, NCCF, Clipping Central and SHR yielded pairs of inputs.

Table 1. Font sizes of headings. Table captions should always be positioned *above* the tables. The final sentence of a table caption should end without a period

Input	ACF NCCF	ACF SHR	ACF CCLIP	NCCF SHR	NCCF CCLIP	SHR CLIP
Results	Bad	Normal	Normal	Worst	Bad	Best

Mixed couples all calculation methods F0 yielded different results. The worst decision was level classifier at NCCF versus SHR. The best classifier had a distinctive character input methods in SHR and Clipping Central, which was to be expected, because these two methods are less prone to adverse extract F0 from unvoiced parts of the signal. Best and worst results are shown in Figure 3 and 4.

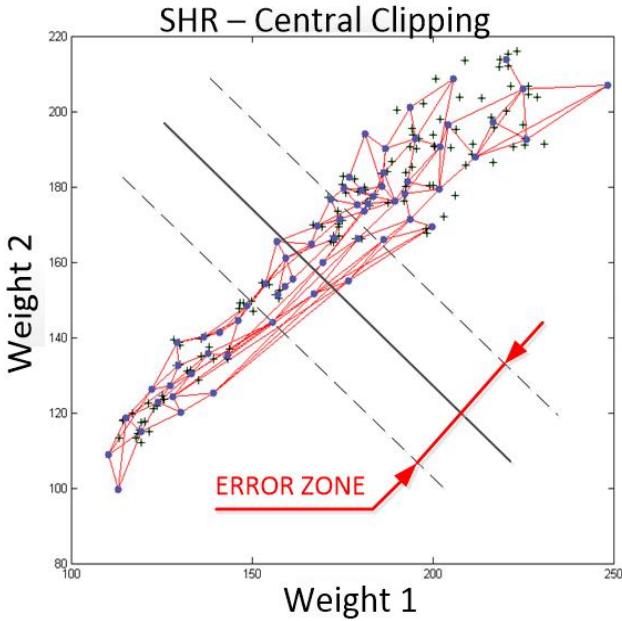


Fig. 3. . Trained self-organizing map with error zones for Subharmonic-to-Harmonic ratio and Central Clipping methods.

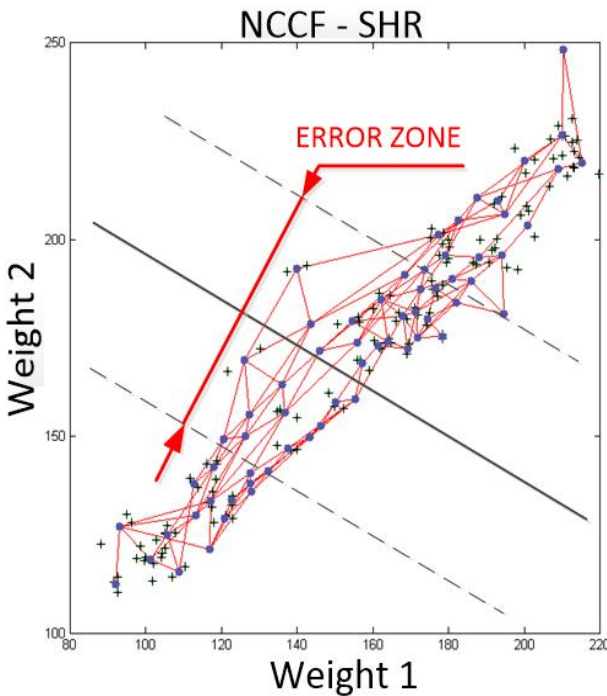


Fig. 4. . Trained self-organizing map with error zones for Normalized cross-correlation function and Subharmonic-to-Harmonic ratio methods.

5 Conclusion

This article focused on the distinctive character classifier. Using different methods for calculating the fundamental frequency of human voice can yield results of different accuracy. Therefore, it is recommended to apply at least two methods for calculating F0. The second perspective is not appropriate to use high dimensional maps for the classification of emotional state. Such classifiers are hardware and time consuming.

Inputs for the two-dimensional neural network were created by combining these four methods. Subharmonic-To-Harmonic Ratio, and Central Clipping methods were considered more reliable than others. It is interesting to note that the classifier has the highest error rate for combined SHR and NCCF. The SOM classifier has the lowest error rate, and thus the best resolving power between normal and stress emotional state, provided SHR and Central Clipping methods had been applied on its inputs. The aim of the article was to determine a suitable combination of the two methods to determine F0 for classifying human emotional state.

References

1. Nicholson, J., Takahashi, K., Nakatsu, R. Emotion Recognition in Speech Using Neural Networks. Springer Verlag: Neural Computing & Applications, Volume 9, Issue 4, 2000, pp. 290-296
2. Partila, P., Voznak, M., Mikulec, M., Zdralek, J. Fundamental Frequency Extraction Method using Central Clipping and its Importance for the Classification of Emotional State. Advances in Electrical and Electronic Engineering, Volume 10, Issue 4, 2012, pp. 270 – 275.
3. Kasi, K., Zahorian, Stephen A., Yet Another Algorithm for Pitch Tracking, 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol.1, 2002, pp.I-361-I-364
4. Picone J. W. Signal modeling techniques in speech recognition. Proceedings of the IEEE. 1993, vol. 81, iss. 9, pp. 1215-1247.
5. Beale M., Howard B., Hudson M. Neural network design. Boulder, Colo: Campus Publ. Service, 2002.
6. Atassi, H.; Smékal, Z.; Esposito, A., Emotion Recognition from Spontaneous Slavic Speech. In Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012). 2012. pp. 389-394. ISBN: 978-1-4673-5185- 0.
7. David Casasent, Sanjay Natarajan, A classifier neural net with complex-valued weights and square-law nonlinearities, Neural Networks, Volume 8, Issue 6, 1995, pp. 989-998, ISSN 0893-6080, [http://dx.doi.org/10.1016/0893-6080\(95\)00008-N](http://dx.doi.org/10.1016/0893-6080(95)00008-N).
8. Tin Lay Nwe, Say Wei Foo and Liyanage C De Silva, Speech emotion recognition using hidden Markov models, Speech Communication, Volume 41, Issue 4, November 2003, pp 603-623, ISSN 0167-6393, [http://dx.doi.org/10.1016/S0167-6393\(03\)00099-2](http://dx.doi.org/10.1016/S0167-6393(03)00099-2).
9. Moataz El Ayadi, Mohamed S. and Kamel, Fakhri Karray, Survey on speech emotion recognition: Features, classification schemes, and databases, Pattern Recognition, Volume 44, Issue 3, March 2011, Pages 572-587, ISSN 0031-3203,
10. Le Yang, Zhongbin Ouyang, Yong Shi, A Modified Clustering Method Based on Self-Organizing Maps and Its Applications, Procedia Computer Science, Volume 9, 2012, pp 1371-1379, ISSN 1877-0509, <http://dx.doi.org/10.1016/j.procs.2012.04.151>.

Study of Electromagnetic Compatibility between Railway Infrastructure and Train Interface

Phan Thi Thanh Thao and Lubomír Ivánek

Department of Electrical Engineering, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
thaoch04_06@yahoo.com, lubomir.ivanek@vsb.cz

Abstract. In railway applications possible interaction between power train and signaling system requires efforts to Electromagnetic Compatibility (EMC) problems. The study is required to verify the processes, procedures and systems of ensuring electromagnetic compatibility between rolling stock and infrastructure in the railway system. The Electromagnetic Interference (EMI) generated by the other systems and to other external EMI sources to which the system may be exposed from the general railway environment and equipment adjacent to the railway infrastructure. And the impact of the infrastructure on the electromagnetic emissions radiated by a railway system.

1 Introduction

As technology advances the railway industry is becoming more reliant on sophisticated systems to perform control and monitoring functions. Communication between stations was improved through the developments of telegraph, telephone and radio systems. Hence, the EMI and EMC issues were considered. Failures in compatibility within and between systems may cause errors of railway systems – signaling, control and telecommunication for example – which can result in hazardous situations with an increased risk of harm to passengers and staff.

There are many practical instances when one is interested in the electromagnetic field propagated from an electrified railway as well as the distribution of leakage currents due to train movement. An understanding of these issues is important in the estimation of the EMI risk to sensitive signaling and telecommunication circuits which are to be installed wayside. Interference issues leading to noise, error and sometimes damage of the track side signaling cables and equipments were experienced widely.

The railway infrastructure could be divided in two parts (fig. 1), the rolling stock represents the first one and the railway power supply is the other part. In our study we will focus on the frequency modeling of a railway power supply infrastructure in frequency band of 10 kHz – 30 MHz which is useful to predict the EMC behavior of the railway system. The railway infrastructure operates in a complex and non-homogeneous environment where low power electronics has to function in the same

environment as large voltages and currents from trains. The environment close to the railway tracks is heavily polluted by electromagnetic noise from the railway systems themselves.

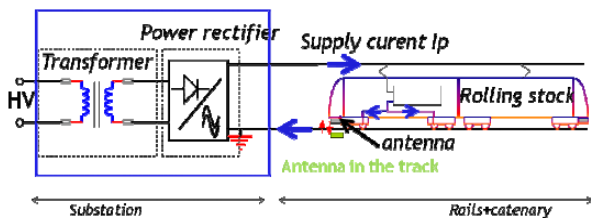


Fig. 1. Simplified diagram of railway infrastructure

In general, the EMC level of a railway system to the outside world is determined by the combined effect of the EM emission sources of the rolling stock and the infrastructure. On the other side, the EM interaction between different parts of a railway system determines the internal EMC level of that system.

As regards emissions of the whole system toward the outside world, in addition to emissions directly radiated from on-board power electronic systems and radiated by the infrastructure. On the other side, it should be noticed that the infrastructure behaves essentially as a very large horizontally-shaped antenna, whose radiation properties strongly depend on a number of parameters, such as the length of the segment used for test execution, the termination loads, discontinuities along the conductors, geometrical asymmetries . . .

- Since a railway line is practically a horizontal wire antenna, conducted EMI propagating along it inevitably excite generating an electromagnetic field. This may pose a serious threat to nearby telecommunication lines, also polluting the frequency spectrum used by radio broadcasters.

- The railway line can also be regarded as a multiconductor transmission line. Therefore it allows the propagation of such EMI signals; since other devices and equipments are connected to the same supply-line, the disturbances generated by in one point may perturb the proper working of other devices, even many kilometers away from the EMI source.

- A railway line is usually not single-track, but it rather presents several parallel tracks, as near as possible one to each other. This leads to electromagnetic couplings between the lines, for example control signals propagating along the rails.

All the considerations discussed above are the motivations for this study, which represents an attempt to model and interpret the dominant phenomena involved in the generation of the emissions radiated by a railway system. In particular, this theoretical analysis is to understand the impact of the electrical and geometrical features of the infrastructure on the radiated emissions.

In the railway system, the most important standard is the so called CENELEC EN 50121. The standards EN 50121 describes the tests to be performed in order to point out EMC problems within a railway system. Together with these tests, limits are imposed, both to conducted and radiated EMI.

2 EMI sources in railway system

The railway is a complex electromagnetic environment and EMC is a significant area of concern. The interference, or noise, mitigation and so electromagnetic compatibility is achieved by addressing both emission and immunity issues, for example suppressing the sources of interference and hardening the potential victims. EMC therefore has both "equipment" and also "environment" elements and is an extension of the functional specification of an equipment, system or installation.

2.1 The railway EMC standards

The standards EN 50121 indicate the methodologies and the limits to apply, relating to the electromagnetic emissions and immunity of railway equipments, vehicles and infrastructures. The series of standards EN 50121 contains the following parts:

- EN 50121-1 Railway applications – Electromagnetic compatibility
Part 1: General
This part describes the characteristics of the railway system which affect electromagnetic compatibility behavior. It specifies a management process for achieving EMC at the interface between the railway infrastructure and trains.
- EN 50121-2 Railway applications – Electromagnetic compatibility
Part 2: Emission of the whole railway system to the outside world
This part sets the emission limits from the railway to the outside world at radio frequencies. Basic statistical tools are introduced in order to get a representative estimation of the measured quantities.
- EN 50121-3-1 Railway applications – Electromagnetic compatibility
Part 3-1: Rolling stock - Train and complete vehicle
This part specifies the emission and immunity requirements for all types of rolling stock. It covers traction stock and trainsets, as well as independent hauled stock.
- EN 50121-3-2 Railway applications – Electromagnetic compatibility
Part 3-2: Rolling stock - Apparatus
This part applies to emission and immunity aspects of EMC for electrical and electronic apparatus intended for use on railway rolling stock.
- EN 50121-4 Railway applications – Electromagnetic compatibility
Part 4: Emission and immunity of the signaling and telecommunications apparatus
This part specifies limits for electromagnetic emission and immunity for signaling and telecommunications apparatus.
- EN 50121-5 Railway applications – Electromagnetic compatibility
Part 5: Emission and immunity of fixed power supply installations and apparatus

This part applies to emission and immunity aspects of EMC for electrical and electronic apparatus and components intended for use on railway fixed installations associated with power supply.

2.2 Description of the method of emission measurement

The purpose of controlling the emissions from electrical and electronic equipment is to protect the general electromagnetic environment and to ensure that radio services and other co-located equipment are able to operate. The control of emissions would allow other equipment intended to be used in the same environment to do so without unacceptable interference. The main aim in this study is to investigate electromagnetic disturbances, radiated or conducted emission from moving trains or from other sources, and to explore whether these emissions could disturb the detector or signaling systems. In order to avoid EMI with high levels of electromagnetic disturbances, measurements are performed.

The basic setup for this test is shown in fig. 2. Here, the electromagnetic field is measured 10 m away from the mechanical axis of a railway track, on which the train to be tested is running. The mechanical centre of the antenna is set at a certain height with respect to the top of the rails. The actual height depends on the test to be performed. Radiated emissions are characterized by measuring the lateral component of the magnetic field over the frequency range 9 kHz – 30 MHz, with the antenna height set in the range 1–2 m. On the other hand, the horizontal and vertical components of the electric field are measured over the frequency range 30 MHz – 1 GHz, with the antenna height set to 3 m. Signals from the antennas are measured in the frequency domain, usually by means of spectrum analyzers. This setup is shared by tests with and without trains. In the latter case, the train is tested under stationary and slow moving conditions (about 60 km/h for main-line trains).

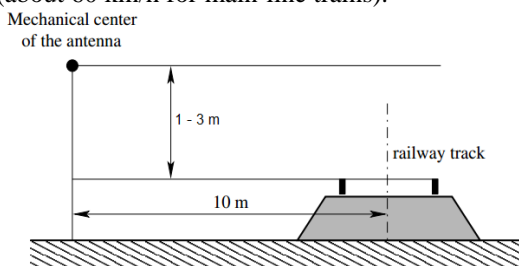


Fig. 2. Measurement set up for frequency domain electromagnetic fields

The measurements protocol and the limits are specified for the frequencies varying from 9 kHz to 1 GHz.

Frequency bands	9 – 150 kHz	0,15 – 30 MHz	30 – 300 MHz	300 MHz – 1 GHz
Bandwidth	200 Hz	9 kHz	120 kHz	120 kHz

To cover the full frequency range, antennas of different design are required. Typical equipment is described below:

Frequency	Type of antenna	Measurement field
9 kHz – 30 MHz	Loop or frame antenna	H
30 MHz – 300 MHz	Biconical dipole	E
300 MHz – 1,0 GHz	Log-periodic antenna (LPA)	E

The standards define among others:

- which component of field, magnetic (up to 30MHz) or electric (above 30 MHz), has to be measured.

- the resolution bandwidth of the spectrum analyzer that must be used for the measurements. Then the limit curve depends on these parameters and is not a constant as shown until 1MHz in the fig. 3.

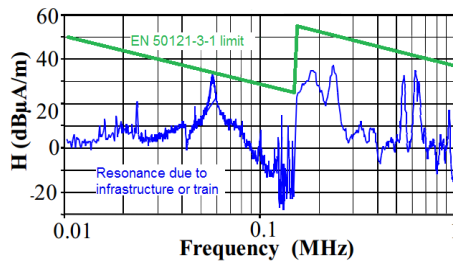


Fig. 3. An example of magnetic field of railway infrastructure

During measurements of the electromagnetic field radiated at 10 m from the railway track, we can observe that, sometimes, for some frequency, the standards limit can be exceeded. An example of measurement is given in fig. 4. The exceeding is characterized by resonance phenomena which appeared for some frequencies of the power supply current.

3 Railway EMC Environment measurement

Sources of electromagnetic disturbance can be naturally occurring, such as lightning, or man-made. Man-made sources may intentionally produce emissions such as bona fide spectrum users or unintentionally such as emissions from equipment, systems or installations. Emissions from equipment may cause spectral pollution, with examples of sources being electrical machines, fluorescent lighting, induction heating, arc welders, inverters, power lines and microprocessor based equipment.

The main threat to equipment within the railway is that due to electric traction (both trains and power transmission). This can cause:

- High electric and magnetic fields at 50 Hz power frequencies and associated harmonics (the AC systems)
- High magnetic fields at both DC and at harmonics on the supply due to the ripple associated with the pulse number of the rectifier in use (DC systems)
- Broadband noise due to contact arcing (pantograph and shoe-gear bounce and conductor gapping).
- Inverter switching harmonics.

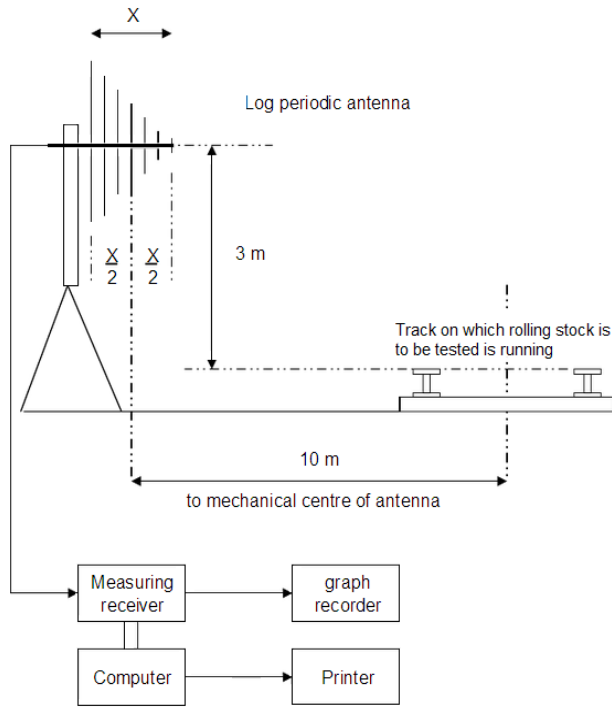


Fig. 4. Measurement of electric field in the 300MHz to 1GHz frequency band

Source of EMC in Railway System

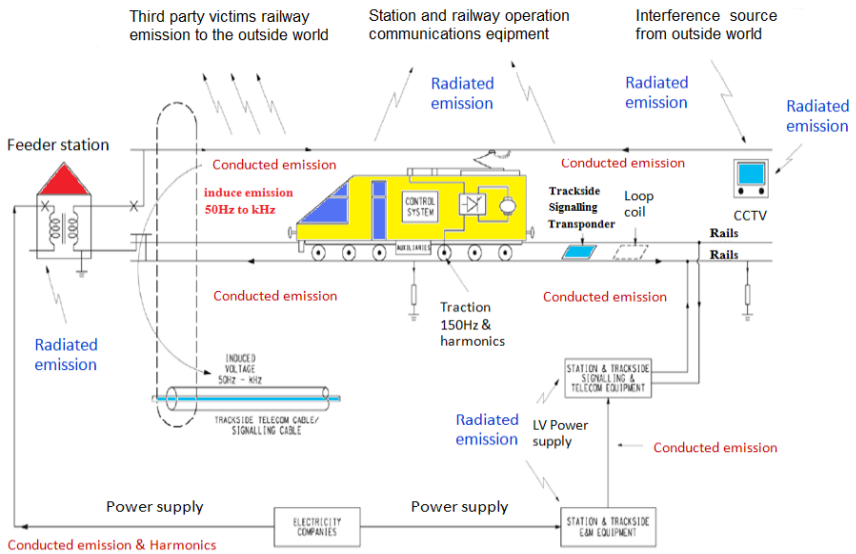


Fig. 5. Radiated electromagnetic disturbances

Other equipments which may be either a source or a victim include:

- Train-borne equipment.
- Signaling systems such as track circuits, axel counters, control equipment.
- Communications systems.

4 Conclusion

Electromagnetic compatibility is achieved when electrical and electronic systems designed for different purposes function effectively and safely, despite interacting in ways that are not necessarily desired. EMC requires the designers of all the systems involved to anticipate this interaction, and collaborate to eliminate or mitigate undesirable consequences.

The research study has helped in understanding the function of railway signaling and detector installations from an EMC and EMI point of view. The knowledge generated will be of assistance in designing new signaling and detector equipment which will have a higher level of reliability, leading to a smaller number of failures and EMC problems. The study has made a contribution towards an understanding of the EMC and EMI characteristics of the signaling and detector infrastructure of the railway system.

References

1. CELENEC EN 50121: Railway applications – Electromagnetic Compatibility
2. Cozza, A.: Railways EMC: Assessment of Infrastructure Impact, Politecnico di Torino, 2005
3. <http://ec.europa.eu/enterprise/sectors/electrical/documents/emc/guidance/>

Recovery of LED Diode Optical Spectrum

Jan Vitásek and Jan Látal

Department of Telecommunications, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jan.vitasek, jan.latal}@vsb.cz

Abstract. The importance and using of power LED diodes increases. White power LED diodes cover a wide spectral range and they are usable in many applications. These LED diodes work in optical systems, the original spectrum of a white power LED diode can be changed by using all kinds of optical elements. This article describes a pursuit of the most faithful recovery of the original spectrum of a white power LED diode. The evaluative criterion is the value of the Correlated Color Temperature of the original white power LED diode compared with the value of the Correlated Color Temperature of the recovered spectrum. This recovery can be used in communication engineering.

1 Introduction

At present the wireless communication technologies are in the process of constant development. The reason is simple: the wireless technologies enable mobility of end users. Except the traditional radio data networks like WiFi, the Free Space Optic (FSO) links are quickly developing and it means that they have a potential of replacing the WiFi networks in the future. The indoor FSO networks have several advantages. First of all, they are not very much affected by the outdoor atmospheric effects. Next, the indoor FSO networks cover the interiors of rooms only, they do not penetrate the walls and therefore they are more resistant to eavesdropping. On the other hand, there are also some disadvantages. A line of sight interruption between a transmitter and a receiver causes a failure in communication. Further on, inside rooms there is some optical noise, which is mainly caused by illuminating light bulbs and fluorescent lamps.

The indoor FSO networks most often use infrared radiation or visible light for communication. There was also an experiment applying ultraviolet radiation, which usage of higher optical powers in comparison with visible or infrared radiation possible. This article deals with research of Visible Light Communication.

2 Visible Light Communication

Visible Light Communication (VLC) is described by standard IEEE 802.15.7 [1]. As the name suggests, this standard uses visible light for communication. Visible light sources can be the white LED diodes that are modulated by changes of the light intensity. The modulation speed must not be perceivable by human eye, it has to be fast enough. Thus VLC provides illumination and data communication at the same time. IEEE 802.15.7 standard offers three physical layer types roughly grouped according to the data rate. The first physical layer type operates from $11.67 \text{ kb}\cdot\text{s}^{-1}$ to $266.6 \text{ kb}\cdot\text{s}^{-1}$, the second type from $1.25 \text{ Mb}\cdot\text{s}^{-1}$ to $96 \text{ Mb}\cdot\text{s}^{-1}$, and the third type from $12 \text{ Mb}\cdot\text{s}^{-1}$ to $96 \text{ Mb}\cdot\text{s}^{-1}$. The first and the second physical layer types use modulation formats consisting of on-off keying (OOK) and of variable pulse-position modulation (VPPM). The third physical layer type uses a modulation format called color shift keying (CSK) [2].

3 LED White Light Creation

There exist two main methods of creating a white light LED diode. The first one uses more color LED diodes, the other one uses a LED diode and suitable luminophor.

In the first group there is a LED diode consisting of three light sources emitting blue (450 nm), green (540 nm) and red (610 nm) light. The correct power setting of these three sources creates a quality white light source [3]. This method is called color mixing.

The second group includes a LED diode emitting blue light that excites yellow phosphor, which leads to phosphorescence and creating yellow light. The merging of the yellow and blue lights creates white light. The yellow phosphor is usually YAG (Yttrium Aluminium Garnet) $\text{Y}_3\text{Al}_5\text{O}_{12}$ [4]. This method is called wavelength conversion.

Both groups have specific modulation limits. Higher optical power is needed in VLC and therefore power LED diodes are used. A white power LED diode consisting of three light sources is supplied by forward current of up to 700 mA [5]. To construct a circuit closer that would be able to switch on and off such a high level of the forward current very quickly is difficult. In the other white LED diode group the modulation rate is dependent on the rate of creating white light, i.e. on the rate of the chemical reaction of yellow phosphor. This chemical reaction cannot be speeded up.

These limits can be overcome by using another solution. A white power LED diode can be suppressed by a notch filter, the suppressed part of the spectrum will be replaced by light from another LED diode that will be modulated. A white power LED diode will emit continuously and due to this the above described problems will cease to exist. The useful data will be transmitted by a narrow spectral LED diode. It will be supplied by lower forward current and therefore it can be modulated faster.

For illumination the white power LED diode Luxeon 5W Star with marking LXHL-LW6C [5] was chosen. The typical forward current of this LED diode is 700 mA, the luminous flux Φ is 120 lm and the color temperature is 5500 K. The emitted spectrum

of this LED diode is shown in Fig. 1. This spectrum was measured by spectrometer USB650.

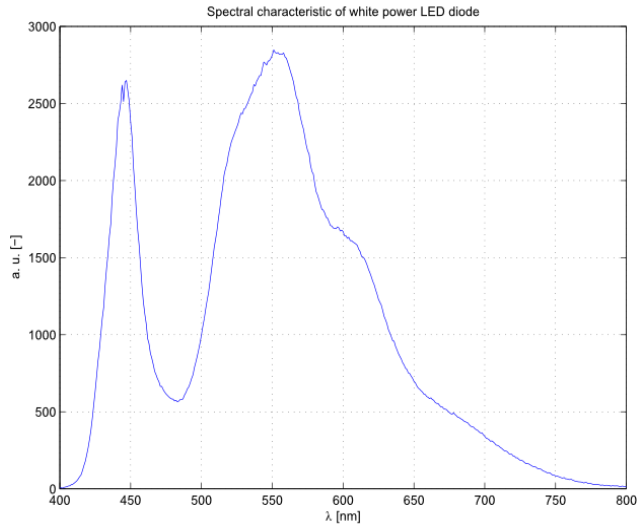


Fig. 1. Spectral characteristics of a white power LED diode

4 Suppression of Part of Optical Spectrum

Notch filter NT67-119 for suppression of part of the spectrum emitted by the white power LED diode was chosen. The features of this notch filter are summarized in Table 1.

Table 1. Features of notch filter

Feature	Value
Central Wavelength	532 nm
FWHM	26.6 ± 2.7 nm
Transmission Wavelength	400–700 nm
Transmission	90 %
Reflection at CWL	99 %
Optical Density	4

The spectral transmission of the notch filter was measured in a laboratory. The spectral transmission is given by the relation

$$T_{\lambda} [\%] = (S_{\lambda} - D_{\lambda}) / (R_{\lambda} - D_{\lambda}) \cdot 100, \quad (1)$$

where S_λ is sample intensity at a wavelength λ , D_λ is dark intensity at wavelength λ and R_λ is reference intensity at wavelength λ . The measured graph of spectral transmission is in Fig. 2.

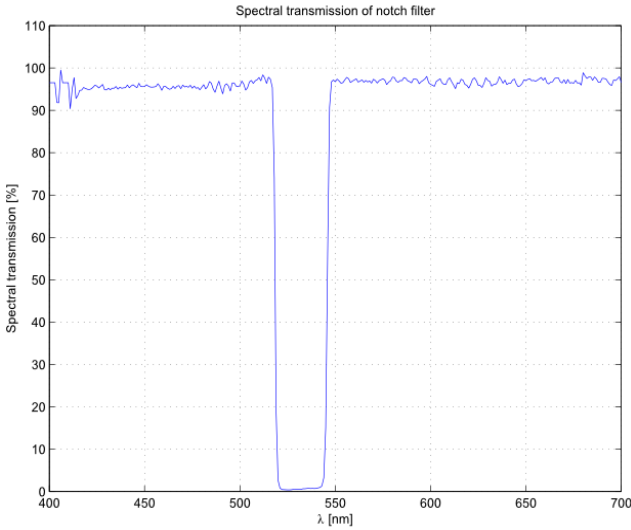


Fig. 2. Spectral transmission of notch filter

The application of the notch filter caused the suppression of a part of the original spectrum. The suppressed spectrum is in Fig. 3.

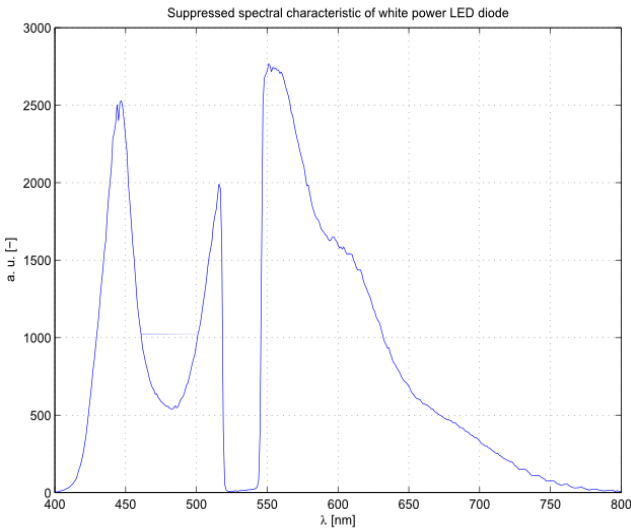


Fig. 3. Suppressed optical spectrum

5 Recovery of Suppressed Optical Spectrum

The article [6] deals with the conducted spectrum recovery by a LED diode and a laser diode. It was proved that the recovery was possible, but it was necessary to find the appropriate LED diode as well as the appropriate forward current. Likewise, it was also proved that the recovery with a laser diode was very difficult and complicated. Several LED diodes that could replace the suppressed part were bought. For each LED diode its spectral characteristic for different forward currents was measured. From each spectral characteristic bottom wavelength λ_b , upper wavelength λ_u as well as wavelength of maximum λ_{max} were set. From the bottom and the upper wavelengths the full width at half maximum FWHM was calculated. Based on the results of the spectral measurements, the power LED diode LXHL-PM01 as the most suitable LED diode for recovery of the suppressed part was chosen. This LED diode has suitable spectral features and it is a 1W LED diode. The notch filter transmits 83.9 % of the optical power of a 5W white power LED diode. It means that a 1W power LED diode can recover the suppressed part of the spectrum.

In measurement of the recovery of the original spectrum spatial merging of two optical beams was needed first. For this purpose a beamsplitter 50/50 % was used. The beamsplitter splits the optical power from two LED diodes in the ratio of 50:50, therefore the correct merging is reached. For better merging a diffuser was also used. The measurement scheme is in Fig. 4.

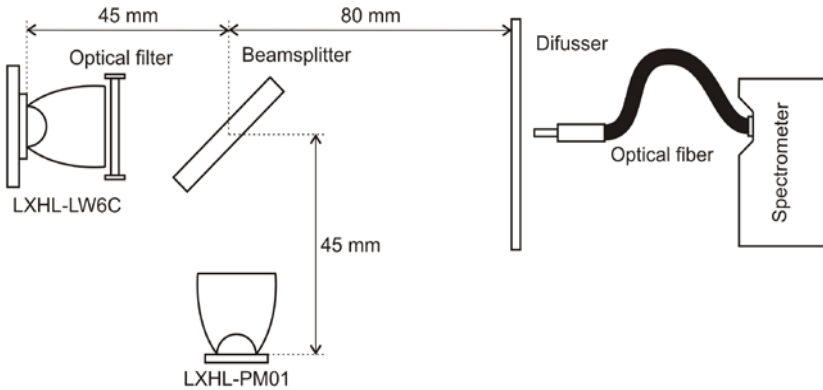


Fig. 4. Measurement scheme of recovery of the original spectrum

The evaluation criterion of recovery of the optical spectrum was the Correlated Color Temperature CCT. At first, the CCT_{org} of the white power LED diode without the notch filter was measured. Further on, the CCT_{rec} after the recovery with the power LED diode LXHL-PM01 was measured. Both these values were compared by calculating the deviation according to the equation (2)

$$\delta [\%] = (CCT_{rec} - CCT_{org}) / CCT_{org} \cdot 100. \quad (2)$$

The following Table 2 summarizes the measured results. The forward current was gradually increased by step 10 mA and the CCT values were recorded. The best result was obtained when power LED diode LXHL-PM01 was supplied by the forward current of 301 mA. The deviation between the original and the recovered white light was only 0.27 %.

Table 2. Results of spectrum recovery in laboratory

U_f [V]	I_f [mA]	CCT_{org}	CCT_{rec}	δ [%]
3.03	281	5541	5490	-0.92
3.05	291	5521	5542	0.38
3.06	301	5545	5560	0.27
3.07	311	5548	5574	0.47
3.08	321	5532	5603	1.28

Fig. 5 shows the original spectrum, the spectrum emitted by LED diode LXHL-PM01 and the recovered spectrum.

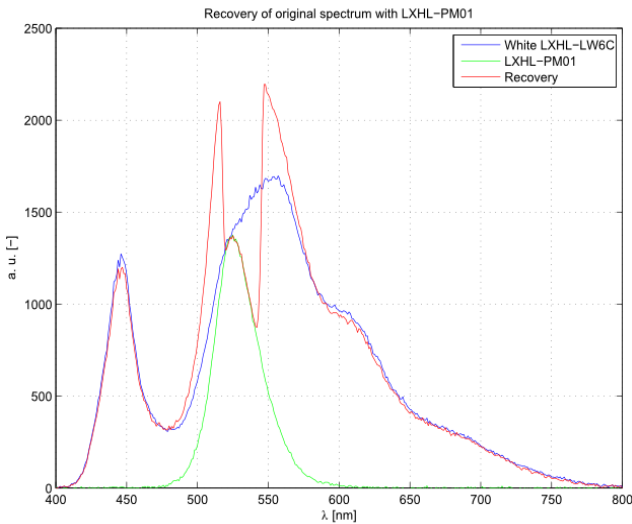


Fig. 5. Recovered spectrum measured in laboratory

6 Conclusion

This article deals with purposeful suppression of the white power LED diode spectrum and the subsequent recovery of the original spectrum. The measurements in the laboratory and the simulation confirmed that recovering of the original spectrum is possible. The evaluation criterion was the Correlated Color Temperature. During the measurement in the laboratory such a level of the forward current was set that the CCT

values were nearly equalized. Looking at the Color Rendering Index it can be seen that the conformity between the original and the recovery values occurred as well.

Acknowledgement

The research described in this article could be carried out thanks to the active support of the Czech Science Foundation GAP108/11/1057 (Synthesis, structure and properties of nanocomposites conducting polymer/phyllsilicate) and the Ministry of Education of the Czech Republic within the projects no. SP2013/88 and SP2013/69 of the VSB-Technical University of Ostrava. This article was also supported by project VG20102015053 (GUARDSENSE - The modern structure of photonic sensors and new innovative principles for intrusion detection systems, integrity and protection of critical infrastructure) and Technology Agency of the Czech Republic TA03020439 (Safety of optical transport network and development of optical components for energy grid and energy transport systems). The research has been partially supported by the project no. CZ.1.07/2.3.00/20.0217 (The Development of Excellence of the Telecommunication Research Team in Relation to International Cooperation) within the frame of the operation programme Education for competitiveness financed by the European Structural Funds and from the state budget of the Czech Republic.

References

- [1] IEEE 802.15 WPAN™ Task Group 7 (TG7): Visible Light Communication. *IEEE* [online]. 2011. Available at: <http://www.ieee802.org/15/pub/TG7.html>.
- [2] ROBERTS, R. D., S. RAJAGOPAL and LIM. IEEE 802.15.7 physical layer summary. In: *IEEE GLOBECOM Workshops (GC Wkshps), 2011*. Houston: IEEE, 2011, pp. 772-776. ISBN 978-1-4673-0039-1. DOI: 10.1109/GLOCOMW.2011.6162558."
- [3] SCHUBERT, E. Fred. *Light-Emitting Diodes*. 2nd edition. Cambridge: Cambridge University Press, 2006. ISBN 978-0-521-86538-8.
- [4] HELD, Gilbert. *Introduction to Light Emitting Diode Technology and Applications*. Boca Raton: CRC Press, 2009. ISBN 978-1-4200-7662-2.
- [5] LUXEON. *Luxeon 5W Star: Technical Data DS30*. 2012. Available at: http://www.roithner-laser.com/datasheets/led_single/hexagonal/lxhl_5w.pdf.
- [6] VITASEK, J., J. LATAL, P. SISKA, P. KOUDELKA and V. VASINEK. Restoration of Optical Spectrum. *Advances in Electrical and Electronic Engineering*. 2013, vol. 11, iss. 3, pp. 227-234. ISSN 1804-3119. Available at: <http://advances.utc.sk/index.php/AEEE/article/view/808>.
- [7] LXHL-PM01. Farnell [online]. 2012. Available at: <http://www.farnell.com/datasheets/1444595.pdf>.



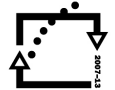
eupean
social fund in the
czech republic



EUROPEAN UNION



MINISTRY OF EDUCATION,
YOUTH AND SPORTS



OP Education
for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT

BIOM Workshop

BIOM workshop – organized as a part of WOFEX 2013 workshop – have been realized as a part of BIOM project (CZ.1.07/2.3.00/20.0073). This project is co-financed by ESF and Czech state budget.



Acknowledgement

Following papers were supported by the Bio-Inspired Methods: research, development and knowledge transfer project, reg. no. CZ.1.07/2.3.00/20.0073, co-financed by ESF and state budget of the Czech Republic:

- Petr Berek, Michal Prílepok, and Václav Snášel. *Similarity detection of Image using Vector quantization and Compression*
- Pavel Bobrov, Alexander Frolov, Dušan Húsek, and Jaroslav Tintěra. *Source localization for EEG patterns relevant to motor imagery BCI control*
- Pavel Dohnálek. *Common Tensor Discriminant Analysis: A Massively Parallel Implementation with Application to Human EEG Wave Recognition*
- Jaroslav Hořejší and Adam Kašpar. *Microsoft Kinect and persons with disabilities*
- Tomáš Jeżowicz, Miloš Kudělka, Jan Platoš, and Václav Snášel. *Visualization of Large Graphs Using GPU Computing*
- Jonáš Krutil, Miloš Kudělka, and Václav Snášel. *Visualization of data using Fibonacci spiral*
- Michal Prílepok, Ibrahim Salem Jahan, and Václav Snášel. *Similarity in EEG data using Lempel-Ziv Complexity*
- Daniel Robenek, Jan Platoš, and Václav Snášel. *Introduction to de novo DNA assembly algorithms*
- Lukáš Zaorálek and Tomáš Buriánek. *Stock price prediction based on particle swarm optimization*

Similarity detection of Image using Vector quantization and Compression

Petr Berek, Michal Prílepok, and Václav Snášel

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{petr.berek, michal.prilepok, vaclav.snasel}@vsb.cz

Abstract. In this paper we show method for similarity detection of image. This method is based on compression algorithm and normalized compression distance (NCD). We show how to prepare image for comparison two images by NCD. NCD needs two original files and one file created by two original files as concatenation. There is question what is concatenation of images. In this paper we describe how to create concatenation of images for using NCD method.

Keywords: NCD, vector quantization, image collection, compression, LZW

1 Introduction

A lot of method for image similarity detection exist. Some method can comparison of pixels in same coordinations. Advanced methods can use some properties or object detection (faces, shapes, etc.) and compare these properties or object. In next section we have described bases of our method. There are quaternions for colour representation, vector quantization for extract blocks of pixels in image and NCD for comparison of two images. In section 3 we describe test collection and settings our experiment. In section 4 we show our results and example similar images searched by our method.

2 State of the Art

2.1 Quaternions

The system of quaternions was put forward in 1843 by W.R. Hamilton (1805-1865). Quaternions were historically the first example of a hypercomplex system, arising from attempts to find a generalization of complex numbers. Complex numbers are depicted geometrically by points in the plane and operations on them correspond to the simplest geometric transformations of the plane. We can use quaternion for colour representation. Quaternion has scalar and vector part. We have a color point $C = (r, g, b)$. We can use quaternion vector part for representation of colour. Vector part is $V(q) = ri + gj + bk$. Instantiated of colours in RGB[17] form we give quaternions vector space[7].

2.2 Similarity of the objects

Data compression can be used for measurement similarity of various objects. There are many data compression algorithm [15] for small text file, images, music files and other type objects.

The main property in the similarity is a measurement of the distance between two objects. The ideal situation is when this distance is a metric [18]. The distance is formally defined as $D : X \times X \rightarrow R$ [5, 13]. The metric is a distance which satisfy three conditions for all $x, y, z \in X$:

1. $D(x, y) = 0$ if and only if $x = y$
2. $D(x, y) = D(y, x)$
3. $D(x, z) \leq D(x, y) + D(y, z)$

The condition 1 is called identity, condition 2 is called symmetry and condition three is the triangle inequality. This definition is valid for any metric, e.g. Euclidean Distance, but the application of this principle into document or data similarity is much complicated.

The basic ideas were suggested and defined in related works by Li et al. [13], and Cilibrasi and Vitanyi . They defined the so-called Normalized Information Distance (NID) [5].

Unfortunately, the Kolmogorov complexity function is non-computable. But Li et al. and Cilibrasi reformulated this problem into a computable form using the replacement of the Kolmogorov complexity by using data compression [5, 13]. The metric developed from their work is a Normalized Compression Distance.

2.3 Normalized Compression Distance

The Normalized Compression Distance (NCD) is based on Kolmogorov complexity. It makes use of standard compressors in order to approximate Kolmogorov complexity. The NCD has been used for text retrieval [11], text clustering, plagiarism detection [4], music clustering [6, 10], music style modeling [8], automatic construction of the phylogeny tree based on whole mitochondrial genomes [12], the automatic construction of language trees [3, 13], and the automatic evaluation of machine translations[19].

The NCD is a mathematical way for measuring the similarity of objects. Measuring of similarity is realized by the help of compression where repeating parts are suppressed by compression. NCD may be used for comparison of different objects, such as images, music, texts or gene sequences. NCD has requirements to compressor. The compressor meets the condition $C(x) = C(xx)$ within logarithmic bounds [16]. We may use NCD for detection of plagiarism and visual data extraction[20, 5].

The resulting rate of probability distance is calculated by the following formula:

$$NCD = \frac{C(xy) - \min(C(x), C(y))}{\max(C(x), C(y))}$$

Where:

- $C(x)$ is the length of compression of x .
- $C(xy)$ is the length of compression concatenation of x and y .
- $\min\{x, y\}$ is the minimum of values x and y .
- $\max\{x, y\}$ is the maximum of values x and y .

The NCD value is in the interval $0 \leq NCD(x, y) \leq 1 + \epsilon$. If $NCD(x, y) = 0$, then files x and y are equal. They have the highest difference when the result value of $NCD(x, y) = 1 + \epsilon$. The constant ϵ describes the inefficiency of the used compressor. The NCD is not a metric. It is an approximation of the NID. The computation of the NCD is very efficient because we do not need to create the output itself. We compute only the size of the output. A study of the efficient implementation of the compression algorithms may be found in [21].

2.4 Similar images detection

We can use NCD for image similarity detection. But we need translate some properties from image into text. With above in text we can use text description image by tag. But we need properties from image. Is possible to create 1D string from 2D image? This process is called as linearization. But can 1D string represent 2D image? This approach is described in [3].

3 Experiment

3.1 Experiment setup

We use vector quantization [9] method for linearization 2D image into 1D string. We divided image into small pieces created by 8x8 pixels. We used quaternion for colour representation [2] of each pixel. So, we created blocks of 8x8 quaternions and codebook was created by these blocks [14]. We looked for coefficient C for relation between codebook compression and image quality after vector quantization. In experiment on collection described below we founded C as $\frac{2}{3}$. Algorithm for block clustering is showed on 1. In our experiment we use own implementation of LZW [22] compression.

3.2 Collection

We used collection from web[1]. There are over 850 images. In collection on web images are in resolution 786x576 pixels. But for our experiment we changed resolution at 314x235 pixels. In collection there are a lot of directories divide images into group (Animals, Flowers, Foliage, Textures, Fruits, Landscapes, Winter, Man mode, Shadows)

```

CB is codebook
|Cs|=size of CB //count of blocks in CB
|Cn|=size of CB //count of blocks in CB
while(Cn > Cs x C)
{
    Find minimal distance between all pair of blocks
    B and NB are blocks with minimal distance between them
    create new block AB as average NB and B and save AB into CB
    delete NB and B from CB
    |Cn|=|Cn|-1
}

```

Fig. 1: Clustering algorithm for quantization

3.3 Saving image for using NCD

LZW method needs string data, we needed to save image as a string. In first we tried to save only indexes from codebook. For each original block we looked for the nearest block in codebook after vector quantization and save index from codebook. Results of this approach were very poor. In next we saved all blocks in codebook and indexes. Each quaternion in each block we saved as string we saved vector part (i.e. RGB parts) as string. We gave text file for compression.

4 Results

Table 1 shows values given by NCD method. We created result in directory stated inside table. Results created from these directories were the best and results are most presentable. On figures below (2, 3) we show example of similar images. First image is origin image too. Search algorithm finds same image as the nearest image of original image. Each example has five images.

Directory	Min value	Max value	Average value
Animals	0,9263	1,1028	0,968
Flowers	0,8376	1,0499	0,8787
Foliage	0,9179	1,0896	0,9719
Textures	0,9151	1,1021	1,0069

Table 1: Values given by NCD



Fig. 2: Example of similarity detection on images in textures directory

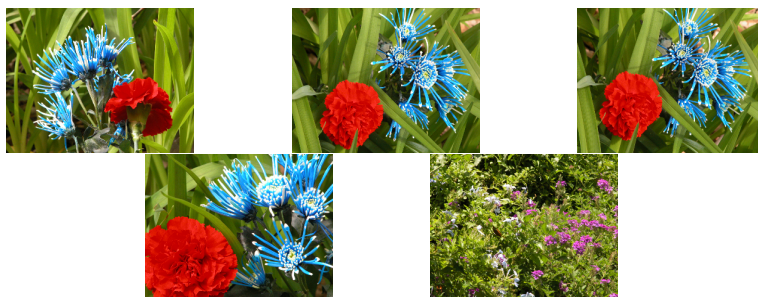


Fig. 3: Example of similarity detection on images in flowers directory

5 Conclusion

In this paper we showed method for image similarity detection based on vector quantization and NCD method. In example we showed this method is functional and this method have practical use. In next paper we want to use another larger collection of image and we want to try functionality of this method for larger collection.

Acknowledgment

This work was partially supported by the Grant Agency of the Czech Republic under grant no. P202/11/P142, and by the Grant of SGS No. SP2013/70, VŠB - Technical University of Ostrava, Czech Republic, and was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and by the Bio-Inspired Methods: research, development and knowledge transfer project, reg. no. CZ.1.07/2.3.00/20.0073 funded by Operational Programme Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic.

References

1. <http://tabby.vision.mcgill.ca/html/welcome.html>. [Online; accessed 11.4.2013].

2. J. Angulo. Computational color imaging. chapter Structure Tensor of Colour Quaternion Image Representations for Invariant Feature Extraction, pages 91–100. Springer-Verlag, Berlin, Heidelberg, 2009.
3. D. Benedetto, E. Caglioti, and V. Loretto. Language trees and zipping. *Physical Review Letters*, 88(4):487021–487024, 2002. cited By (since 1996)145.
4. X. Chen, B. Francia, M. Li, B. McKinnon, and A. Seker. Shared information and program plagiarism detection. *IEEE Transactions on Information Theory*, 50(7):1545–1551, 2004.
5. R. Cilibrasi and P. M. B. Vitányi. Clustering by compression. *IEEE Transactions on Information Theory*, 51(4):1523–1545, 2005.
6. R. Cilibrasi, P. Vitányi, and R. de Wolf. Algorithmic clustering of music based on string compression. *Computer Music Journal*, 28(4):49–67, 2004. cited By (since 1996)76.
7. L. Dorst, D. Fontijne, and S. Mann. *Geometric algebra for computer science : an object-oriented approach to geometry*. Morgan Kaufmann series in computer graphics. Elsevier San Francisco, Amsterdam, 2007.
8. S. Dubnov, G. Assayag, O. Lartillot, and G. Bejerano. Using machine-learning methods for musical style modeling. *Computer*, 36(10):73–80, 2003. cited By (since 1996)25.
9. A. Gersho and R. M. Gray. *Vector quantization and signal compression*. Kluwer Academic Publishers, Norwell, MA, USA, 1991.
10. A. González-Pardo, A. Granados, D. Camacho, and F. De Borja Rodríguez. Influence of music representation on compression-based clustering. 2010. cited By (since 1996)0.
11. A. Granados. Analysis and study on text representation to improve the accuracy of the normalized compression distance. *AI Commun.*, 25(4):381–384, 2012.
12. M. Li, J. Badger, X. Chen, S. Kwong, P. Kearney, and H. Zhang. An information-based sequence distance and its application to whole mitochondrial genome phylogeny. *Bioinformatics*, 17(2):149–154, 2001. cited By (since 1996)274.
13. M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitányi. The similarity metric. *IEEE Transactions on Information Theory*, 50(12):3250–3264, 2004.
14. C.-Y. C. T.-C. Lu. *A Survey of VQ Codebook Generation*. Journal of Information Hiding and Multimedia Signal Processing, 2010.
15. J. Platos, V. Snášel, and E. El-Qawasmeh. Compression of small text files. *Advanced Engineering Informatics*, 22(3):410–417, 2008.
16. D. Sculley and C. Brodley. Compression and machine learning: A new perspective on feature space vectors. pages 332–341, 2006. cited By (since 1996)17.
17. S. Süsstrunk, R. Buckley, and S. Swen. Standard rgb color spaces. In *In The Seventh Color Imaging Conference: Color Science, Systems, and Applications*, pages 127–134, 1999.
18. A. Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977. cited By (since 1996)1968.
19. J. J. Väyrynen, T. Tapiovaara, K. Kettunen, and M. Dobrinkat. Normalized compression distance as an automatic MT evaluation metric. In *Proceedings of MT 25 years on*, 21–22 Nov 2009 Cranfield, UK, to appear.
20. P. M. B. Vitányi. Universal similarity. *CoRR*, abs/cs/0504089, 2005.
21. J. Walder, M. Krátký, R. Baca, J. Platos, and V. Snášel. Fast decoding algorithms for variable-lengths codes. *Inf. Sci.*, 183(1):66–91, 2012.
22. J. Ziv and A. Lempel. Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, 24(5):530–536, 1978.

Source localization for EEG patterns relevant to motor imagery BCI control

Pavel Bobrov^{1,2}, Alexander Frolov^{1,2}, Dušan Húsek^{1,3}, and Jaroslav Tintěra⁴

¹IT4I Vysoká škola báňská – Technická universita Ostrava,
17. listopadu 15/2172 708 33 Ostrava Poruba, Czech Republic
p-bobrov@yandex.ru

²Institute for Higher Nervous Activity and Neurophysiology of Russian Academy of Sciences, Butlerova 5a, Moscow Russian Federation
aafrolov@mail.ru

³Institute of Computer Science, Academy of Sciences of the Czech Republic,
Pod Vodárenskou Věží 2, Prague 8, Czech Republic
dusan@cs.cas.cz

⁴Institute for Clinical and Experimental Medicine,
Vídaňská 1958/9, Praha, Czech Republic,
tinti57@gmail.com

Abstract. This work concerns spatial localization of sources of EEG patterns the most specific for control of the motor imagery based BCI. In our previous work we have shown that performance of Bayesian BCI classifier can be drastically improved by extraction of the most relevant independent components of the EEG signal. This paper presents the results of spatial localization of electrical brain activity sources which activity is reflected by the extracted components. The localization was performed by solving the inverse problem in EEG source localization, using individual finite-element head models. The sources were located in central sulcus (Brodmann area 3a), in the superior regions of post- and precentral gyri, and supplementary motor cortex.

1 Introduction

Brain-computer interfacing has become an established technology during the last years. The most common brain computer interfaces are those which are based on classification of electroencephalogram (EEG) patterns specific to performing different motor imagery tasks, i.e. tasks involving a kinesthetic imagining of a particular movement. We have shown earlier that a basic Bayesian classifier applied to raw EEG data can distinguish between different palm contraction imagining and relaxation, with accuracy comparable to that of the more sophisticated methods [1]. It has been shown also that Independent Component Analysis (ICA) can be used to both extract EEG patterns the most relevant to the tasks performed during experiment and to enhance the classifier performance [2, 3]. The first goal of this work is to test how the ICA-enhanced Bayesian classifier can handle classification of four tasks, namely, relaxation, left and right palm movement imagining, and left foot movement imaging. The second goal is to extract EEG patterns which are the most relevant to performing of the tasks and to perform their spatial localization.

2 Experiment protocol and data acquisition

Sixteen subjects participated in the study (4 male and 4 female). None of the subjects had neurological diseases reported. All experiments were carried out in the Institute of Computer Science of the Academy of Sciences of Czech Republic.

The subjects have been training to perform motor imagery tasks for 10 days, one experimental session a day. The subjects had to imagine slow contraction of his left palm, to imagine slow contraction of his right palm, to imagine slow pressing a pedal with his left foot, and to relax. Each experimental session consisted of 10 blocks, each block consisted of 4 different instructions (relax or perform a certain motor imagery task) presented in pseudorandom order. The sequence of instructions was generated anew at the beginning of each session. The subject had to perform each task for 10 seconds during each block. There was a 5 second preparatory period preceding the task. In total, session lasted for 10 minutes.

The data were acquired using 40-channel setup of g.Tec USBamp digitizers with g.Tec active electrodes (g.Tec Medical Engineering GmbH, Austria), 256 samples per second, 24 bit per channel. The electrodes were placed at Fz, F4, F3, Fcz, Fc3, Fc4, Fc5, Fc6, Cz, C2, C1, C4, C3, C6, C5, Cpz, Cp1, Cp2, Cp3, Cp4, Cp5, Cp6, Pz, P2, P1, P4, P3, P6, P5, Poz, Po3, Po4, Fp1, Oz, O1, O2, Po7, Po8, F7, P8, Fc7, Tp8, T7, T8, Tp7, Fc8, P7, and F8 positions specified by the extended 10-20 International system. The data were band-pass filtered with built-in Butterworth filter in 5-30 Hz pass band.

At the day of the experiment each subject took part in fMRI study after the BCI training session. During the study the subjects had to perform the same tasks as during the BCI control. In addition to functional MRI, a T1-weighted anatomical scan using MP-RAGE sequence (224 slices, TR=2300 ms, TE=4.64 ms, FOV=256×256, slice thickness 1 mm, voxel size of 1×1×1 mm, PAT=2, sagittal orientation) was acquired for each subject.

3 Data processing

First, independent components the most relevant to performing of the experiment tasks were extracted. Second, the sources which activity was reflected by the extracted components were localized by solving the inverse problem in EEG source localization.

The data recorded during the experiments was decomposed into the sum of independent components using RunICA algorithm implemented in EEGLab toolbox for MATLAB [4]. It was shown recently that choice of ICA algorithm does not affect distribution weights and activity of the components which have an established physiological interpretation [5].

To select independent components the most significant for BCI performance Cohen's kappa (κ) index was computed in dependence on the number N_{cmp} of ICA components used for mental state classifying. Cohen's kappa was computed based on the confusion

matrix elements which were estimated by cross-validation of the session data with the Bayesian classifier.

For each N_{cmp} we found the optimal combination of components providing the highest κ . Since the total number of possible component combinations is extremely large (2^{48}) we used exhaustive search to find the optimal combination of only for $N_{cmp} = 3$ components. To find the optimal combination of components for $N_{cmp} > 3$ we used a "greedy" algorithm which added components one by one starting from the optimal combination of 3 components. At each step a component was added to the optimal component combination found at the previous step, so that extended combination provided the highest κ .

With respect to the EEG analysis, it is reasonable to assume that independent sources of electrical brain activity recorded at the head surface are resulting from current dipoles distributed over the neocortex. Our experiments confirm that for the sources the most relevant to the BCI performance the distribution of electrical potential over the head produced by each of these sources could be actually interpreted in terms of electrical field produced by single current dipole. Thus, for each of such sources its localization was searched in a single dipole approximation. In other words, position and orientation of a single current dipole were searched which provided maximal matching between patterns of EEG distribution on the head surface given by ICA and by a dipole. The pattern of EEG distribution for dipole with given position and orientation was computed by solving the direct EEG problem. It was solved by the finite element method (FEM) which allows taking into account individual geometry of the brain and its covers.

To generate the FEM meshes from the MRI data, MR images were segmented into five sub-regions: white matter, gray matter, cerebrospinal fluid (CSF), skull and scalp. The segmentation of the different tissues within the head was made by means of SPM8 New Segmentation Tool. To construct the FE models of the whole head the FEM mesh generation was performed using tetrahedral elements with inner-node spacing of 2 mm. Thus, the total number of nodes amounted to about 1.5 millions. Electrical conductivities were assigned to the tissues segmented in accordance with each tissue type: 0.14 S/m for white matter, 0.33 S/m for gray matter, 1.79 S/m for CSF, 0.0132 S/m for skull, and 0.35 S/m for scalp [6]. To solve the EEG forward problems, the FEM mesh along with electrical conductivities were imported into the commercial software ANSYS (ANSYS, Inc., PA, USA).

4 Results

The dependency of κ on the number of independent components selected is shown in Fig. 1. It can be seen from the figure that κ depends on number of the components not monotonically and reaches maximum when some ICA components are discarded. However the classifying accuracy for 3 components is very close to that obtained for optimal combination of components providing maximal κ . Thus the obtained combinations of 3 components can be considered to represent EEG patterns the most specific to performing of the motor imagery tasks.

The components of the optimal triple happened to be not identical for all experimental days and all subjects but some of them appeared frequently. The distributions of contribution into signal on EEG electrodes of four components which appeared most often are demonstrated in Fig 2. Three of these components $\mu 1$, $\mu 2$ and $\mu 3$) demonstrate very well exposed Event Related De-synchronization (ERD) of mu-rhythm. For the component denoted $\mu 1$ mu-rhythm is suppressed during the left hand motor imagery. Respectively, for the components $\mu 2$ and $\mu 3$ mu-rhythm is suppressed during the right hand and feet motor imagery and their foci are presumably above the areas presenting right hand and feet. The fourth component (denoted β) we ascribed to the activity of supplementary motor areas (SMA). First, it is located on the midline surface of the hemispheres and hence their activity can produce the focus corresponding to the topoplot of this component. Second, this component demonstrated ERD in the spectral band 10 - 40 Hz during performing of all motor imagery tasks.

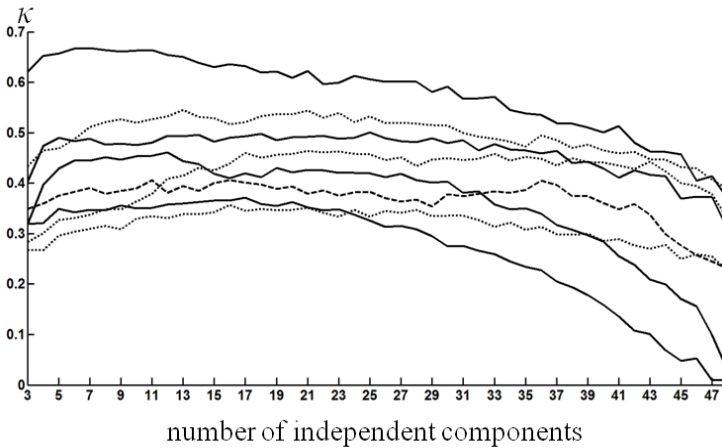


Fig. 1. Index κ of BCI control accuracy in dependence on the number of ICA components used for mental states classifying. Each curve represents the data for each individual subject obtained for the last day of BCI training.

Dipole positions obtained for components $\mu 1$ and $\mu 2$ happened to be very close to the sensorimotor “hand areas“, similarly to the results obtained earlier [7]. Dipole position obtained for component $\mu 3$ is close to the sensorimotor “feet areas“, in the superior regions of post- and precentral gyri and for component β slightly more anterior, i.e. close to SMA. However, the positions of dipoles corresponding to $\mu 1$ and $\mu 2$ happened to be a little deeper than areas of fMRI activity near “hand areas“.

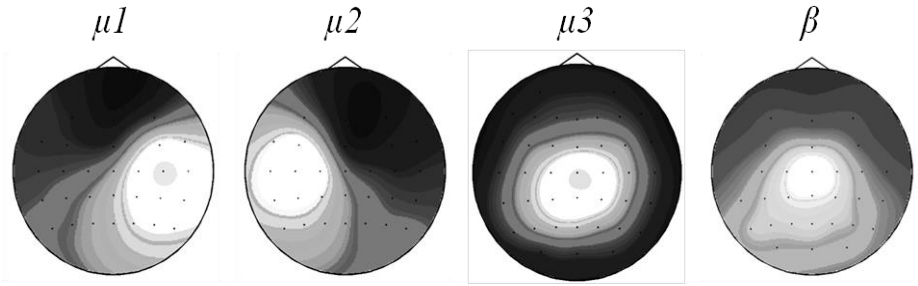


Fig. 2. Topoplots of four sources of electrical brain activity the most relevant to BCI performance which appeared most often.

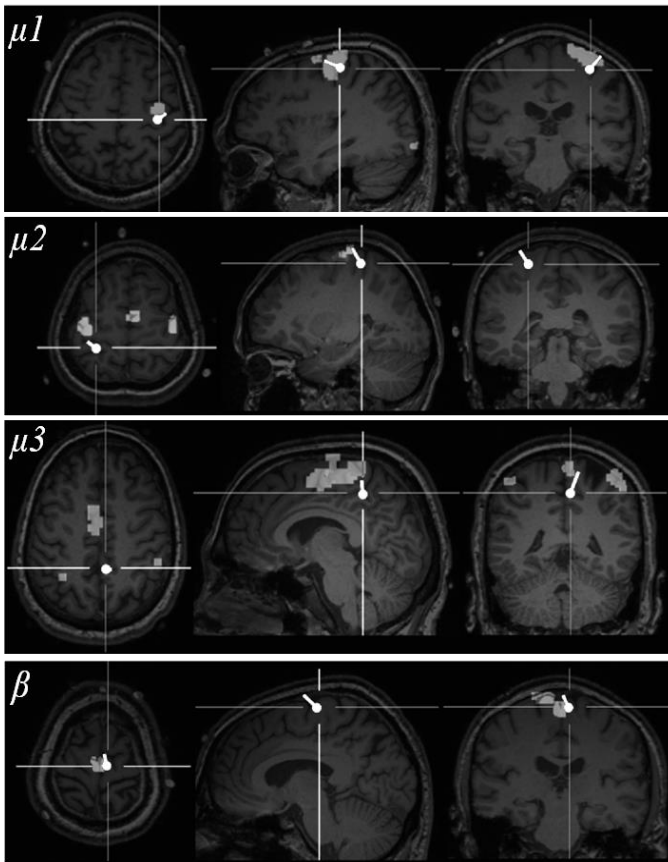


Fig. 3. The results of source localization are shown for subject 1.

In Fig. 3 the $\mu 1$ and β sources are shown along with voxels in which BOLD response was significantly higher during left hand MI than during relaxation, $\mu 2$ is shown along with voxels in which BOLD response was significantly higher during right hand MI than during relaxation, and $\mu 3$ is shown along with voxels in which BOLD response was significantly higher during foot MI than during relaxation.

5 Conclusion

BCI training and ICA analysis allowed for extraction of the EEG patterns the most relevant for motor imagery task execution. The patterns can be considered as resulting from time-varying activity of compact brain areas, and the activity can be successfully approximated by varying amplitude of a single dipolar source. We have obtained a precise localization of the dipolar sources of the relevant patterns with a tool which utilizes individual head geometry of each subject. This spatial information is intended to be used for validation of source localization based BCI which is being designed by our group.

Acknowledgments

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and by the Bio-Inspired Methods: research, development and knowledge transfer project, reg. no. CZ.1.07/2.3.00/20.0073 funded by Operational Programme Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic.

References

1. Frolov A., Husek D., Bobrov P. (2011). Comparison of four classification methods for brain computer interface *Neural Network World*, 21(2) 101-192.
2. Kachenoura A., Albera L., Senhadji L., Comon P. (2008). ICA: a potential tool for BCI systems. *Signal Processing Magazine, IEEE*, 25(1), 57-68.
3. Bobrov P., Frolov A., Husek D. (2012). Brain Computer Interface Enhancement by Independent Component Analysis. *Proceedings of the Third International Conference on Intelligent Human Computer Interaction (IHCI 2011), Advances in Intelligent Systems and Computing*, 179, 51-60.
4. Delorme A., Makeig S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 134, 9-21.
5. Delorme A., Palmer J., Onton J., Oostenveld R., Makeig S. (2012). Independent EEG sources are dipolar. *PLoS one*, 7(2), e30135.
6. Wolters C.H., Anwander A., Tricoche X., Weinstein D., Koch M.A., MacLeod R.S. (2006). Influence of tissue conductivity anisotropy on EEG/MEG field and return current computation in a realistic head model: a simulation and visualization study using high-resolution finite element modeling. *NeuroImage*, 30:813-26.
7. Frolov A., Husek D., Bobrov P., Korshakov A., Chernikova L., Konovalov R., Mokienko O. (2012). Sources of EEG activity most relevant to performance of brain-computer interface based on motor imagery. *Neural Network World*, 22(1), 21-37

Common Tensor Discriminant Analysis: A Massively Parallel Implementation with Application to Human EEG Wave Recognition

Pavel Dohnálek

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
`pavel.dohnalek@vsb.cz`

Abstract. In this paper, a massively parallel implementation of Common Tensor Discriminant Analysis (CTDA) is presented with applications to human brainwave pattern recognition. The implementation, accelerated by the NVIDIA CUDA technology, is shown to be 11.49x faster than the original MATLAB version. Before processing by CTDA, the data is segmented by a sliding window and converted into the time-frequency domain by the continuous wavelet transform (CWT).

1 Introduction

Brain computer interfaces (BCI) are still considered to be the holy grail of human-computer interaction and as such are given a significant amount of attention and research efforts. With the variety of applications ranging from simple interaction with personal computers to aiding systems for disabled people to extensive military use, BCIs also have a tremendous potential in business. While most papers and researches focus on increasing the recognition accuracy of various brainwave classification techniques, it is just as necessary to be able to perform the classification as fast as possible with the minimum cost of the hardware used. In this paper, a massively parallel implementation of the CTDA preprocessing algorithm is presented and compared to its original MATLAB implementation in terms of computational performance.

2 Signal preprocessing

To improve recognition accuracy and sometimes computation speed, data is usually preprocessed prior to classification. The classic approach to preprocessing is signal decomposition, data factorization, feature extraction and selection etc. One of the most fundamental concepts in the computer science is the Fast Fourier Transform (FFT) [3]. It is universal in signal processing but can also be used, among other things, to compress image and audio files, solve differential equations and price stock options [7]. Other methods use the principles of signal

decomposition and approaches similar to FFT. Here, wavelet transforms, one of which is also used for data preprocessing in this paper, should be mentioned first with its wide use in different areas [5]. The Hilbert-Huang Transform (HHT) can be mentioned as a member of the class of algorithms that decompose a signal into so-called intrinsic mode functions (IMF) and obtain instantaneous frequency data. It is designed to work well for non-stationary and non-linear data. In contrast to other common transforms like the FFT, the HHT is more of an algorithm (an empirical approach) that can be applied to a data set rather than a theoretical tool [6].

3 CTDA

Originally presented by Zhao et al. in [9], CTDA has been successfully applied in BCI systems achieving recognition accuracy of up to 95%. An overall description of the CTDA algorithm is provided in this section. It explains the tensor representation of an EEG signal, formally describes the CTDA feature extraction process and points out where exactly is the algorithm suitable for massive parallelism.

3.1 EEG Tensor Representation

Suppose that there is an electroencephalographic (EEG) epoch represented as a $N_c \times N_t$ matrix $\mathbf{X} \in \mathbb{R}^{N_c \times N_t}$ where N_c is the number of channels used for registration and N_t is the number of samples. Then, in order to capture the time-frequency structure of the signal the epoch \mathbf{X} can be transformed into tensor $\mathcal{X} \in \mathbb{R}^{N_c \times N_f \times N_t}$ using Continuous Wavelet Transform (CWT) of its channels, i.e.

$$\mathcal{X}(c, :, :) = \text{CWT}(\mathbf{X}(c, :)), c = 1, \dots, N_c \quad (1)$$

when MATLAB notation is followed. Here N_f is the number of scales for which wavelet coefficients are computed.

Notion of a sample covariance matrix can be also generalized for tensor representation by defining a covariance tensor \mathcal{R} as $\frac{1}{N_t} [\mathcal{X} \circ \mathcal{X}^T | 3] \in \mathbb{R}^{N_c \times N_f \times N_f \times N_c}$ with \mathcal{X}^T denoting a tensor from $\mathbb{R}^{N_f \times N_c}$ for which $\mathcal{X}_{pqt}^T = \mathcal{X}_{qpt}$ (note the permutation of \mathcal{R} dimensions). The elements of covariance tensor are

$$\mathcal{R}_{i,m,n,j} = \frac{1}{N_t} \sum_x \mathcal{X}_{i,m,x} \mathcal{X}_{j,n,x}. \quad (2)$$

The notion of covariance tensor plays an important role in CTDA.

3.2 CTDA Problem Formulation and Feature Extraction

In general there is a set of several EEG records, $\mathbf{X}^l, l = 1, \dots, L$, attributed to different classes. In practice \mathbf{X}^l are obtained during a training session and are readily available. In order to extract features using CTDA records \mathbf{X}^l are

split into epochs $\mathbf{X}^{l,i}$, $i = 1, \dots, N_l$, whose lengths are equal and determine how often the classifier output is generated. The epochs are then transformed into tensors $\mathcal{X}^{l,i}$ as described in the previous subsection. For each of the tensors $\mathcal{X}^{l,i}$ its covariance tensor $\mathcal{R}^{l,i}$ is computed. These covariance tensors can be used to estimate the "typical" covariance tensor for the l -th class as

$$\mathcal{R}^l = \frac{1}{N_l} \sum_i \mathcal{R}^{l,i}. \quad (3)$$

Given the estimates of covariance tensors of each class the CTDA problem is to find matrices $\mathbf{W}_{l,1}$ and $\mathbf{W}_{l,2}$, $l = 1, \dots, L$, (two matrices for each class) projecting the signal corresponding to performing the l -th class on the first and second modes respectively so that

$$\begin{aligned} \mathcal{R}^l \times_1 (\mathbf{W}^{l,1})^T \times_2 (\mathbf{W}^{l,2})^T \times_3 (\mathbf{W}^{l,2})^T \times_4 \\ \times_4 (\mathbf{W}^{l,1})^T = \mathcal{D}^l, \end{aligned} \quad (4)$$

$$\begin{aligned} \left(\sum_l \mathcal{R}^l \right) \times_1 (\mathbf{W}^{l,1})^T \times_2 (\mathbf{W}^{l,2})^T \times_3 (\mathbf{W}^{l,2})^T \times_4 \\ \times_4 (\mathbf{W}^{l,1})^T = \mathcal{J}^l \end{aligned} \quad (5)$$

and

$$\begin{aligned} \mathcal{D}_{ijkl}^l &= d_i d_j \delta_{il} \delta_{jk}, \\ \mathcal{J}_{ijkl}^l &= \delta_{il} \delta_{jk} \end{aligned} \quad (6)$$

Once the matrices $\mathbf{W}^{l,1}$ and $\mathbf{W}^{l,2}$ are obtained for each l , feature extraction can be performed as follows. Let \mathcal{Y} be an EEG epoch represented as a tensor. First, \mathcal{Y} is projected using matrices \mathbf{W} for each class yielding L tensors

$$\mathcal{Z}^l = \mathcal{Y} \times_1 (\mathbf{W}^{l,1})^T \times_2 (\mathbf{W}^{l,2})^T \quad (7)$$

After that, vectors v_l , $l = 1, L$, are calculated as $v_l^T = \text{diag}(\text{mat}_3(\mathcal{Z}^l) \text{mat}_3(\mathcal{Z}^l)^T)$. These vectors are concatenated so that $v^T = v_1^T \dots v_L^T$ and the resulting vector is log-transformed $v = \log(v)$ component-wise. The vector v is then the feature vector corresponding to the epoch \mathcal{Y} . Complete formal specification of the CTDA process can be found in [9]. In the testing phase, equation 7 is computed for every test epoch, thus generating feature vectors to classify. The classifier itself was based on Euclidean distance.

3.3 Parallel implementation

Generally, all the equations mentioned above can be computed in parallel. The most time consuming parts lie in tensor to matrix multiplications which can be

transformed into matrix to matrix multiplication with respect to appropriate tensor unfolding. cuBLAS routines (GEMM) [8] were used in our implementation to process such multiplications. Next, many hidden operations were done in parallel as well, e.g. sum of elements on a matrix diagonal, square roots of matrix elements etc. Because the CTDA is based on the Singular Value Decomposition (SVD) [4], the CULA parallel library [1] was used to compute this matrix decomposition. During CTDA computation, unfolding is performed many times and it is also highly parallelized in the implementation.

Definition 1. (Unfolding): *The Mode- n unfolding of tensor $\underline{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is denoted by $Y_{(n)}$ and arranges the Mode- n fibers into columns of a matrix. More specifically, a tensor element (i_1, i_2, \dots, i_N) maps onto a matrix element (i_n, j) , where*

$$j = 1 + \sum_{p \neq n} (i_p - 1) J_p, \quad \text{with} \quad (8)$$

$$J_p = \begin{cases} 1, & \text{if } p = 1 \text{ or if } p = 2 \text{ and } n = 1, \\ \prod_{m \neq n}^{p-1} I_m, & \text{otherwise.} \end{cases}$$

Template function was used according to the best practices in CUDA. This allows adjusting the compiled code with respect to target unfolding mode, e.g. Mode-1 \rightarrow 2 or Mode-1 \rightarrow 3. The proposed method maximizes CUDA register utilization, avoids time-consuming modulo operations and minimizes access into global memory space.

4 Experiments

The following section describes the dataset used and the overall experimental setting, summarizes the experiments performed and presents the resulting computation times as well as accuracy. The experiments were performed on a personal computer composed of 32 GB RAM, 8-core AMD FX-8150 3.6Ghz and Tesla C2075 GPU. The used software equipment was 64b Windows 8 Professional operating system, CUDA 5.0 and latest versions of cuBLAS, CULA and MATLAB.

4.1 Data collection

During the experiment the subject had to either relax, imagine slow contraction of his left or right palm or imagine slowly pressing his foot against a pedal, thus having 4 tasks to perform. He was sitting in a comfortable chair in front of a computer monitor with a circle drawn in the center of the screen the purpose of which was to fixate the subject's gaze. The beginning of the task to perform was clued by changing the color of one of the four gray arrows drawn around the

Table 1. Performance comparison between CUDA and MATLAB implementation

Phase	Measurement	Implementation		Speed-up
		CUDA	MATLAB	
Training	Feature vectors	451	451	11.49x
	Iterations	30	9	
	Total Time (s)	350.41	1207.76	
	Iteration time (s)	11.68	134.20	
Testing	Feature vectors	1802	1802	5.68x
	Total Time (s)	45.161	250.39	
	Iteration time (ms)	24.48	138.95	

circle into green which is a common practice when acquiring EEG data [2]. The session was split into 8 periods, each period containing all tasks presented at a random order. The subject had to keep performing each task for 10 seconds so the experiment lasted 320 seconds.

4.2 Data arrangement

Overall, 144384 time samples were recorded from each of the sensors, some of which were not part of the experiment and could be omitted. The samples were windowed into 1 second long epochs (256 values) with the window shift of 0.25s. This created 2553 labelled epochs, 451 of which were used to train CTDA, the rest for testing. After the CWT processing, the training tensor's dimensions were $48 \times 15 \times 115456$, the testing tensor's $48 \times 15 \times 461312$.

4.3 Results

Given the tensor dimensions, computing CTDA and generating feature vectors from the testing set was time-consuming enough for it to make sense to measure the experiments in seconds. Table 1 shows the measured results. The first row represents the number of iterations required for the CTDA algorithm to converge. Although the same number of iterations for both implementations is expected, the difference can be explained through the different implementations of SVD in CULA and MATLAB. Still, the computation times are generally significantly lower. The speed-up of computing each iteration in the training phase reached the factor of 11.49 while the speed-up of generating test vectors for classification was lower at 5.68x. The lower speed-up is easily explained by the fact that generating test vectors is a much simpler operation of simply extracting the features from test signals, thus weakening the CUDA parallelism advantage and, at the same time, increasing the significance of unavoidable memory transfer delays. Still, more than 5 times faster processing can be considered satisfactory as this increase in performance allows real-time classification.

5 Conclusion

This paper reviewed the performance of a parallel implementation of the CTDA preprocessing technique in applications to EEG wave recognition. It was shown that a simple change of the programming paradigm can greatly benefit the computation speed of both the CTDA algorithm itself and extracting features for classification. The main advantage of the paradigm is that it is widely available even in low-cost personal computers and should be made available for mobile devices in the near future as well. The main disadvantage, on the other hand, is more complicated practical programming. Possible directions for future work include decreasing the computational complexity of the algorithm to be computable on low-power and mobile devices, allowing for the algorithm to be implemented in easily portable and possibly even wearable EEG solutions.

Acknowledgment

This article has been elaborated in the framework of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 funded by Structural Funds of the European Union and state budget of the Czech Republic. The work is partially supported by Grant of SGS No. SP2013/70, VŠB - Technical University of Ostrava, Czech Republic. This work was also supported by the Bio-Inspired Methods: research, development and knowledge transfer project, reg. no. CZ.1.07/2.3.00/20.0073 funded by Operational Programme Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic.

References

1. CULA Tools: GPU Accelerated Linear Algebra, 2010.
2. Charles Anderson, Elliott Forney, Douglas Hains, and Annamalai Natarajan. Reliable identification of mental tasks using time-embedded eeg and sequential evidence accumulation. *J Neural Eng*, 8(2):025023, 2011.
3. Hubert Cecotti and Axel Graeser. Convolutional Neural Network with embedded Fourier Transform for EEG classification. pages 1–4, December 2008.
4. L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.
5. D.L. Fugal. *Conceptual wavelets in digital signal processing: an in-depth, practical approach for the non-mathematician*. Space & Signals Technical Pub., 2009.
6. N.E. Huang and N.E.H.S.S. Shen. *The Hilbert-Huang Transform and Its Applications*. Interdisciplinary Mathematical Sciences Series. World Scientific, 2005.
7. R.J. Marks. *Handbook of Fourier Analysis & Its Applications*. Oxford University Press, USA, 2009.
8. NVIDIA. *CUBLAS Library User Guide*. NVIDIA, v5.0 edition, October 2012.
9. Qibin Zhao, Liqing Zhang, and A. Cichocki. Multilinear generalization of common spatial pattern. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 525–528, 2009.

Microsoft Kinect and persons with disabilities

Jaroslav Hořejší and Adam Kašpar

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{jaroslav.horejsi, adam.kaspar}@vsb.cz

Abstract. We would like to present hardware tools for persons with disabilities; Microsoft Kinect will be introduced as well as its current use for persons with disabilities, and also its use in the EduKin project. The paper provides essential information about EduKin.

Keywords: Microsoft Kinect, Trackballs, keyboard, Head Pointer, I4Control, persons with mental disabilities, persons with physical disabilities, EduKin, autism

1 Introduction

Today, various hardware tools exist that can be used for computer control. Microsoft Kinect is one of these recent tools, which has been used especially for Xbox game control. However, it has recently been found that this tool can also be used for rehabilitation, both physical and occupational.

We will present the development of Microsoft Kinect and its present use around the world for persons with various disabilities. Furthermore, you will be informed about Microsoft Kinect use in the EduKin project.

2 Hardware aids for persons with disabilities

Currently, the following hardware and software tools are available for computer control, designed for persons with disabilities [1]:

- Trackballs - mice adapted in various ways
- Joysticks - joysticks as a mouse replacement
- SmartNAV4 - EG - used as a mouse replacement and designed to scan head movements
- Keyboards - various coloured keyboards or keyboards with large keys, ergonomic design, etc
- Head pointer - computer control using a head pointer
- I4Control - computer control using the eyes
- TrackerPro - computer cursor control using a reflexive point on the head and a camera
- JetVoice - computer control using the voice

3 Microsoft Kinect

Microsoft Kinect is a device designed to control software applications using the human body. This device has been developed for Microsoft Xbox game control. The following subchapter presents a brief history of Microsoft Kinect.

3.1 History of Microsoft Kinect

The first efforts to design a similar hardware tool date back to about 30 years ago, which is when *the Put-That-There project* was created, managed by Chris Schmandt of the Architecture Machine Group at the Massachusetts Institute of Technology. However, this device needed a room of 5 m x 3.5 m, a screen and a vinyl chair placed 2.5 m in front of the screen. The user had to wear a magnetic cube on their wrist and a microphone on their head. The device could recognize voice and gestures.

Another attempt at creating a similar device was made in 1993 at the same place, called *The Iconic System*. Besides voice and gesture recognition, this system was also able to follow eye movement. Special gloves were used instead of magnetic cubes. This project was managed by David Koonz, Kristinn Thorrison and Carlton Sparrell.

The DreamSpace project was launched at the end of the 90s, managed by Mark Lucente for IBM Research. This project was similar to *Put-That-There* with the difference being that a camera system was used for recognition purposes.

In 2003, Microsoft initiated its own research into *Surface* multi-touch control devices.

In 2005, Nintendo presented a game console called Wii Remote, which was able to determine motion in 3D space, and for this reason Microsoft wanted to design a better device. The Natal project was gradually prepared and then presented in 2010. This is how Microsoft Kinect was born. [15]

3.2 Hardware characteristics of the device

Kinect is 263 mm long, 66 mm deep and 39 mm high. . The Kinect head can be

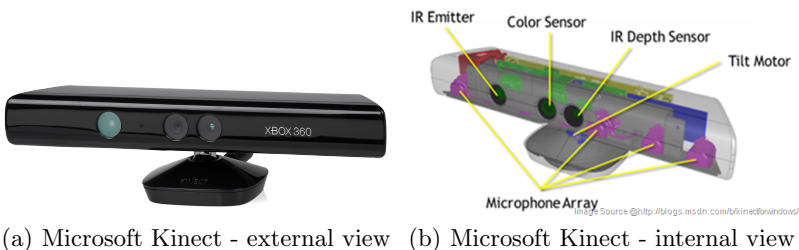


Fig. 1. Picture Microsoftu Kinect - external and internal view

electrically moved. An infrared projector, two cameras, four microphones and ventilation holes are installed inside, see Figs 1(a) and 1(b).

4 Use of Microsoft Kinect for persons with various disabilities

Various Kinect research projects have been carried out worldwide for persons with disabilities, which focus on persons with mental, physical, combined and psychological disabilities. Such research has been carried out particularly in China, in the US and in Germany.

4.1 Use of Microsoft Kinect for rehabilitation exercises

Kinect is used for rehabilitation exercises using a video. When people perform rehabilitation every day, the need to keep them motivated arises. This is a method to keep them motivated, given that the exercise becomes entertaining [2] [3]. In addition, Kinect has been used for balance exercises after a cerebral event using various games [4]. The range of articular movements can also be measured [5].

4.2 Use of Microsoft Kinect for persons with mental disabilities

Research focused on activities that persons with mental disabilities tend to forget has also been done. This pertains particularly to working in restaurants, where these persons forget meal preparation procedures, etc. Microsoft Kinect helps these persons by checking whether the meal is being prepared correctly, and if not, the device reminds the employees accordingly [6].

4.3 Use of Microsoft Kinect for persons with auditory disorders

The Microsoft device can also be used for persons with auditory disorders. Kinect can namely recognize sign language and use this recognition capacity in educational games for persons with auditory disorders [7]. Obviously, this feature can also be used in computer games for persons with auditory disorders [8].

4.4 Use of Microsoft Kinect for persons with vision impairments

Kinect has been used for navigation in unfamiliar buildings [9]. In addition, computer games with Microsoft Kinect can be used for vision improvement [10]. Yet another possibility is to use the device for determining any position on a touch-map [11].

4.5 Autism

Kinect can also be used to become aware of oneself. This can improve social interactions of persons with autism using an artificially created mirror and other virtual objects displayed in the mirror [12]. Furthermore, Kinect can be used to express the feelings of children with autism and of healthy children in computer games. The capacities of children with autism can be evaluated based on the monitored feelings [13].

5 EduKin and a description of the rehabilitation game *Catching objects*

The project is focused on helping children with various learning, movement and other disorders that need more intense care from their teachers and parents. [16] Other games for Kinect are currently being developed as part of the EduKin project. The game called *Catching objects* is described here.

5.1 *Catching objects*

The game *Catching objects* has been implemented by Adam Kašpar in cooperation with the Social Care Institute for Persons with Mental Disabilities. The game should improve locomotive powers of clients. The game is illustrated in Figs. 2(a) 2(b).

Four inclined planes are shown on the screen of the game, and an egg or coin



Fig. 2. Screen of the game *Catching objects*

slides down on them. The task of the player is to catch these objects in a basket or in a piggy bank before they fall to the ground. An appropriate sound is played. The piggy bank or the basket can be controlled using the right or left hand or using both hands at the same time [15].

6 Testing

Application testing is being performed at the Social Care Institute for Persons with Mental Disabilities in Ostrava, where a computer with Kinect installed is available in a special class. Besides basic software, a Windows service has been developed, which is installed for administrative purposes and sends logs at regular intervals. The games will be tested by a select group of children at the Social Care Institute in Ostrava for persons with mental disabilities. Improvements in the success rates of the selected group will be observed during the test. The tests are currently being processed.

7 Conclusion

The purpose of this paper was to summarize the development and present use of Microsoft Kinect for persons with various disabilities. Another purpose was to present the use of Kinect in the EduKin project, including one game that uses Kinect.

References

1. PETIT company (2013) Hardware aids. Available at: <http://www.petit-os.cz> (Accessed: 17 July 2013).
2. A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities, Yao-Jen Chang, Shu-Fang Chen, Jun-Da Huang, National Taiwan Normal University, Taipei, Taiwan
3. Kinerehab: A Kinect-based System for Physical Rehabilitation A Pilot Study for Young Adults with Motor Disabilities, Jun-Da Huang, National Taiwan Normal University, Taipei, Taiwan
4. Development and Evaluation of Low Cost Game-Based Balance Rehabilitation Tool Using the Microsoft Kinect Sensor, Belinda Lange, Chien-Yen Chang, Evan Suma, Bradley Newman, Albert Skip Rizzo, Mark Bolas, 33rd Annual International Conference of the IEEE EMBS, Boston, Massachusetts USA, August 30 - September 3, 2011
5. Chronic Pain Rehabilitation with a Serious Game using Multimodal Input, Christian Schnauer, Thomas Pintaric, Hannes Kaufmann, Interactive Media Systems Group, Institute of Software Technology and Interactive Systems, Vienna University of Technology, Austria, Stephanie Jansen Kosterink, Miriam Vollenbroek-Hutten, Roessingh Research and Development, Enschede, The Netherlands, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede
6. A gesture recognition system to transition autonomously through vocational tasks for individuals with cognitive impairments Yao-Jen Chang - Department of Electronic Engineering, Chung Yuan Christian University, Chung-Li, Taiwan, Shu-Fang Chen - Graduate Institute of Rehabilitation Counseling, National Taiwan Normal University, Taipei, Taiwan, An-Fu Chuang - Association of Employment Rights for People with Disabilities, Taipei, Taiwan

7. American Sign Language Recognition with the Kinect, Zahoor Zafrulla, Thad Starner, Harley Hamilton, Peter Presti, Sch. of Interactive Computing, Georgia Inst. of Technology, Atlanta, GA 30332-0760, Helene Brashear Tin Man Labs, LLC, Austin, Texas
8. Developing a gesture-based game for deaf/mute people Using microsoft kinect Fakhteh Soltani, Fatemeh Eskandari, Shadan Golestan, Department of Computer Engineering, Faculty of Engineering, Arak University, Arak, Iran
9. NAVI A Proof-of-Concept of a Mobile Navigational Aid for Visually Impaired Based on the Microsoft Kinect, Michael Zllner, Stephan Huber, Hans-Christian Jetter, and Harald Reiterer, University of Konstanz, Germany
10. Using Xbox 360 Kinect Games on Enhancing Visual Performance Skills on Institutionalized Older Adults with Wheelchairs, I-Tsun Chiang, Jong-Chang Tsai, Shang-Ti Chen, National Changhua University of Education, Changhua, Taiwan, 2012 Fourth IEEE International Conference On Digital Game And Intelligent Toy Enhanced Learning
11. Kin touch: Understanding How Visually Impaired People Explore Tactile Maps, Anke Brock, Delphine Picard, Samuel Lebaz, Christophe Jouffrais, Bernard Oriola, Philippe Truillet, Universit Toulouse and CNRS, CHI 2012, May 510, 2012, Austin, Texas, USA
12. A kinect-based augmented Reality System for individuals with autism spectrum disorders, Xavier Casas, Gerardo Herrera, Inmaculada Coma, Marcos Fernandez Institute of Robotics, University of Valencia, Valencia, Spain
13. Mimicking Expressiveness Of Movements By Autistic Children In Game Play, Daniel Tetteroo, Azadeh Shirzad, Mariana Serras Pereira, Matthijs Zwinderman, Duy Le, Emilia Barakova, Eindhoven University of Technology, Eindhoven, The Netherlands
14. Computers and Handicaped, Jaroslav Hoej, VB-Technical university of Ostrava, WOFEX 2011, Ostrava
15. Framework for Working with Gestures based on Kinect Technology, Adam Kapar, VB-Technical university of Ostrava, thesis 2013, Ostrava
16. EduKin(2013) EduKin project. Available at: <http://www.edukin.eu> (Accessed: 20 July 2013).

Visualization of Large Graphs Using GPU Computing

Tomáš Ježowicz, Miloš Kudělka, Jan Platoš, and Václav Snášel

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{tomas.jezowicz.st, milos.kudelka, jan.platos, vaclav.snasel}@vsb.cz

Abstract. Graphs may be used to visualize relationships between objects. Relations are represented by edges and objects are called nodes. When graph is drawn, one can easily see and understand the basic structure of data. Many different applications can be found in social network analysis, computer networks, scientific literature analysis, etc. However drawing large graphs (thousands or a millions of nodes), is still challenging problem. There exist many different algorithms for drawing graphs. Each algorithm has specific behavior and different applications and limits. Some algorithms are focused on quality while others are more suitable for large graphs. This paper aims to speed up the computation using GPU, so larger graphs can be visualized in acceptable time, or visualization can be done even in real-time.

Keywords: Fruchterman-Reingold, large graphs, GPU computing, fast graph visualization

1 Introduction

Fast visualization of large graphs is in general a very important task. Graph drawing provides a quick overview of data internal organization. Large networks, having thousands of vertices and edges, can be found in many different areas [1]. For example social networks, web pages connected by links, computer networks, or protein similarity [8].

Different layouts may correspond to one particular graph. Some of the aesthetical criteria for measuring quality of a graph can be to minimize edge crossings or distribute vertices in the frame [4]. The computation time usually increases with the range of aesthetical criteria required. For large graphs it is worth to either compromise quality or try to speed up the computation with some parallel approach.

Today's graphic cards (GPUs) are very efficient at parallel computations, like manipulating computer graphics. General Purpose GPU utilizes graphic processing unit to perform non-graphic computations. The main difference of GPU/CPU is that GPUs can create thousands of light-weight threads and run them in parallel. CUDA (Computed Unified Device Architecture) has become a

popular parallel computing platform in recent years. It provides heterogeneous programming model, which enables user to use both C/C++ and low or higher level GPU API.

The paper is organized as follows. Section II presents related work. In Section III we briefly describe the GPU computing, graph layout problem and Fruchterman-Reingold algorithm. In the Section IV, we characterize the proposed algorithm that speeds up the current implementation. Finally in Section V we describe the experiments with both random graphs and real-world data.

2 Related work

CUDA is already being used for graph related computations. In [6] was showed fast solutions of BFS (Breadth First Search), SSSP (Single Source Shortest Paths), and APSP (All Pairs Shortest Paths) on large graphs with usage of CUDA. Frishman [3] created brand new algorithm by combining both Kamada-Kawai and Fruchterman-Reingold approaches. New parallel analysis of the multipole method of graph drawing was described in [5]. Another interesting paper is [9] where authors were speeding up NodeXL plugin for Microsoft Excel. Computing layouts that are both aesthetically pleasing and fits well into the sequence of drawings of the evolving graph was introduced in [2]. In our work, we focused on general undirected graphs. One of the well known algorithms for graph drawing is Fruchterman-Reingold [4]. We decided to speed up this algorithm, because even the classical version of this algorithm is an order of magnitude faster than Kamada-Kawai [7].

3 Fruchterman-Reingold

The key principles of the algorithm states, that vertices connected by an edge should be drawn near each other, but not too close. Also connected vertices attract each other.

```

area ← W * L {width and length of the frame}
G ← (V, E) {random initial positions for the vertices}
k ←  $\sqrt{\text{area}/|V|}$ 
function  $F_a(z)$  ← { return  $x^2/k$  }
function  $F_r(z)$  ← { return  $k^2/z$  }

```

```

for  $i \leftarrow 1$  to iterations do
  {calculate repulsive forces}
  for all  $v \in V$  do
    {each vertex has two vectors: .pos and .disp}
     $v.\text{disp} \leftarrow 0$ 
    for all  $u \in V$  do
      if  $u \neq v$  then
        {  $\Delta$  is short hand for the difference}

```

```

    { vector between positions of the two vertices }
     $\Delta \leftarrow v.pos - u.pos$ 
     $v.disp \leftarrow v.disp - (\Delta/|\Delta|) * F_r(|\Delta|)$ 
  end if
end for
end for

{calculate attractive forces}
for all  $e \in E$  do
  {each edge is an ordered pair of vertices .v and .u}
   $\Delta \leftarrow e.v.pos - e.u.pos$ 
   $e.v.disp \leftarrow e.v.disp - (\Delta/|\Delta|) * F_a(|\Delta|)$ 
   $e.u.disp \leftarrow e.u.disp - (\Delta/|\Delta|) * F_a(|\Delta|)$ 
end for

{limit the max displacement to the temperature t}
{and then prevent from being displaced outside frame}
for all  $v \in V$  do
   $v.pos \leftarrow v.pos + (v.disp/|v.disp|) * \min(v.disp, t)$ 
   $v.pos.x \leftarrow \min(W/2, \max(-W/2, v.pos.x))$ 
   $v.pos.y \leftarrow \min(L/2, \max(-L/2, v.pos.y))$ 
end for
end for

```

4 Algorithm Speedup

As showed above, the Fruchterman-Reingold basically consists of three independent parts. The most computationally intensive part is calculation of repulsive forces, which requires $\Theta(N^2)$ and can be easily parallelized. The remaining parts works with edges ($\Theta(E)$) and again with all vertices ($\Theta(N)$). We decided to speed up only calculation of repulsive forces.

4.1 Repulsive forces

Our basic idea of speeding up the Fruchterman-Reingold algorithm is to speed up only the slowest part, where each vertex visits each vertex. The rest of computation is performed by CPU, because it is more appropriate. So we divided the whole calculation of repulsive forces to four parts.

1. Copy data to GPU.
2. Run main kernel
3. Run fix kernel
4. Free GPU resources.

Data consists of read-only 2D (x, y) coordinates that are randomly placed in the beginning. The input data size is $2N$, where N is number of vertices. Main kernel is more interesting, since it creates temporary results or final results

(if input size fits to one block). To make it better for parallel processing, we slightly modified the mathematical equation in the first part, so each thread then calculates these equations separately and avoids the expensive root square operation.

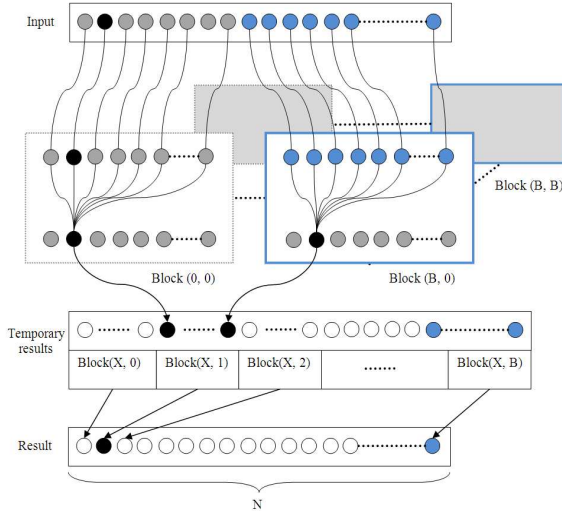


Fig. 1. Principle of algorithm speedup

Figure 1 illustrates the basic idea. The input data is divided to multiple blocks, to fit to input size (N). The x coordinates of each block then covered whole input. In other words if block dimension is 512, then input (of size N) is divided by 512. Number of blocks is B . The y coordinate in the grid of blocks (grid y) was used to identify the output id. One block in the grid reads its input. Block with $id \leftarrow (x = 1, y = 2)$ will read second, and third part of the input (separated by dimension of the block). Once each block has its data in shared memory, block iterates through y data and each thread in x data computes its part. Then simple reduction is performed.

5 Performance Tests

We randomly created different graphs with different count of vertices and edges. The table 1 shows measured results. The test showed, that our solution is approximately 70-80 times faster than current implementation, while achieving the same output. With growing number of edges both CPU and GPU versions tended to ran slower. It was expected, because our implementation has speeded up only the part computing vertices. Edges are computed separately, so the same amount of work must be performed in both versions.

Table 1. Algorithm Speedup with different $|V|$ and $|E|$

$ V $	$ E $						
	0	1K	10K	100K	200K	500K	1M
500	4.8	4.8	3.7	1.9	1.8	1.5	1.1
1K	18.2	17.8	16.4	6.6	4.2	2.4	1.8
5K	115.6	112.1	95.5	58.4	55.1	54.4	20.5
10K	91.6	88.7	78.1	74.9	64.9	66.7	23.7
50K	89.3	90.9	91.1	73.9	84.9	69.5	69.5
100K	90.8	76.93	71.6	90.5	74.4	74.1	68.7
200K	78.9	71.7	74.3	74.5	83.1	74.0	77.17

6 Real-world Data Experiment

For our experiment we took well known DBLP database¹ and all records in GECCO conference². Each author is represented by a vertex and each edge means, that connected authors published at least one publication. The graph contains 5674 vertices and 9557 edges. We added one vertex which connects all the others together, because the graph contains 965 components, because algorithm is not designed for disconnected graphs. After graph coordinates were computed, the previously added vertex was removed. The number of iterations were experimentally set to 2000. Total time on CPU was 4757.14 seconds. When GPU was used, the total time was 47.82 seconds. So the total speedup, that we achieved was 99.50.

7 Hardware configuration and parameters of the algorithm

In our experiments we used a server with AMD Opteron processors @GHz with Linux CentOS, Tesla C2050 with 448 CUDA cores and 3GB RAM. The parameters of Fruchterman-Reingold algorithm were defined as follows. The number of threads per block is limited to amount of shared memory used per block. Our implementation requires $5 \times BLOCKSIZE \times sizeof(float)$ shared memory. The current GPUs can allocate up to 16KB shared memory, which is just

¹ <http://dblp.uni-trier.de/>

² <http://www.sigevo.org/gecco-2013/>

fine for $BLOCKSIZE = 512$. For completeness, all configurable parameters used in tests are: $speeddivisor = 800$, $areamultiplier = 10000$, $speed = 1$, $area = 10000$ and $gravity = 10$.

8 Conclusion and Future Work

This paper describes the speedup of Fruchterman-Reingold force-based graph layout algorithm, while achieving the same results. Our solution can be easily parallelized even on multiple GPUs (by sending whole coordinates to each GPU and then dividing by y grid dimension). This approach appears to be promising both on artificially created data and real-world data. Our future work is focused on computing coordinates for millions of nodes on multiple GPUs or keeping each iteration fast enough to visualize it in real-time on even larger graphs.

Acknowledgment

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), by the Bio-Inspired Methods: research, development and knowledge transfer project, reg. no. CZ.1.07/2.3.00/20.0073 funded by Operational Programme Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic, and by SGS, VSB-Technical University of Ostrava, under the grant no. SP2013/70.

References

1. Batagelj, V., Mrvar, A.: Pajek-program for large network analysis. *Connections*, 1998, 21.2: pp. 47–57
2. Frishman, Yaniv, and Ayellet Tal.: Online dynamic graph drawing. *Visualization and Computer Graphics*, IEEE Transactions on 14.4 (2008): 727-740.
3. Frishman, Yaniv, and Ayellet Tal.: Multi-level graph layout on the GPU. *Visualization and Computer Graphics*, IEEE Transactions on 13.6 (2007): 1310-1319.
4. Fruchterman, Thomas MJ, and Edward M. Reingold.: Graph drawing by forcedirected placement. *Software: Practice and experience* 21.11 (1991): 1129-1164.
5. Godiyal, Apeksha, et al.: Rapid multipole graph drawing on the gpu. *Graph Drawing*. Springer Berlin Heidelberg, 2009.
6. Harish, Pawan, and P. J. Narayanan.: Accelerating large graph algorithms on the GPU using CUDA. *High Performance ComputingHiPC 2007*. Springer Berlin Heidelberg, 2007. 197-208.
7. Kamada, T., Kawai, S.: An algorithm for drawing general undirected graphs. *Information processing letters*, 1989, 31.1: 7-15.
8. Novosád, Tomáš, et al.: Proxima: Protein similarity algorithm. *Nature & Biologically Inspired Computing*, 2009. NaBIC 2009. World Congress on. IEEE, 2009.
9. Sharma, Puneet, et al.: Speeding up network layout and centrality measures for social computing goals. *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer Berlin Heidelberg, 2011. 244-251.

Visualization of data using Fibonacci spiral

Jonáš Krutil, Miloš Kudělka, and Václav Snášel

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{kru178, milos.kudelka, vaclav.snasel}@vsb.cz

Abstract. Visualization of data is one of the most important challenges which data analyst have to go through it. Drawing your data in a simple, obvious and visually appealing way, is very important to your readers to understand, what you want to show them. In addition, drawing your data helps you to understand the data better. As you can see, the data visualization is essential for both readers and data analyst. The key question is, what is the right model for the data visualization. We think, there is no clear answer. It depends on many factors. In this paper, we attempt to introduce our approach, how we visualize data using Fibonacci spiral. Our work is focused on sorting all the nodes in a graph into golden ratio spiral, to help our users easily understand relations in the social networks. We want to provide intuitive way, how to draw relatively small graphs. We expect better user orientation without any additional help or explanation.

1 Introduction

Our primary target is extend visualization options in an online analysis tool Forcoa.NET [1]. This web browser tool is focused on the analysis and visualization of the co-authorship relationship based on the intensity and topic of joint publications from the field of computer science. The social network relationships represented by nodes and edges, Forcoa.NET team introduced two parameters - the retention and stability [2]. These relationships are calculated by forgetting function, based on Forgetting Curve [1] and are visualize in a graph.

A graph is an abstract structure contains objects, where some pairs of objects are connected to each other by links. We call these links as edges and connected objects as nodes. If we want to visualize our data into the graph, there are many possibilities how to draw nodes and their edges. Commonly used are arc diagram, area grouping, centralized ring, circled globe, elliptical implosion, flow chart, radial convergence, radial implosion, scaling circles, trees, and etc [3]. Segmented radial convergence could be the most similar to our approach.

In our approach we are using a golden spiral which is a logarithmic spiral whose growth factor is ϕ , the golden ratio. That is, a golden spiral gets wider (or further from its origin) by a factor of ϕ for every quarter turn it makes [4]. For an approximation the golden spiral we use Fibonacci spiral using quarter-circle arcs inscribed in squares of integer Fibonacci-number.

The main idea is drawing graph nodes and their edges clearly and intuitively into golden spiral, to easily visualize social network relationships. We start drawing the golden spiral from a central node which is represented by one selected author. Then we build the co-authors network around him. We want to help our users to understand, without bigger explanation, who are the most important co-authors in the

network and how close they are to selected author (central node). If the graph contains any subgraphs, the users should identify them without major problems.

2 Related work

Visualization of data from social networks is a popular subject with rich history. Visualizing Social Networks by L. C. Freeman [5] is a paper which reviews the long history of visualization. Mr. Freeman begins with ad-hoc rules for placing the nodes in hand-draw images and continue through development examples of systematic procedures. In Mr. Freeman's article we can see how color, position, shape or size, can be used to encode information.

In this paper, we describe how a node position could be significant, in compare to central node. How a node importance could be represents in a network by his centrality is shown in [6]. Article introduced the notion of derivatives of centrality as a metric that relates all the nodes from social and other scale-free networks in a graph. Their analysis can be applied to network analysis in general.

Another visualization system for exploring large-scale social networks is Vizster. Vizster builds node-edge layouts to explore connectivity in large graph structures, supports visual search and social analysis [7]. Automatically can identify and visualize community structures.

Visualization of graphs is the subject of a survey "Graph visualization and navigation in information visualization". The survey shows good examples of data visualization and provide detailed look into problematic of graphs drawing [8].

3 Our research

The goal is provide to our users more transparent view on relationships in social networks and extend visualize possibilities of analysis tool Forcoa.NET. The project Forcoa.NET provide a view on relationships of the authors publishing in journals and on computer science conferences.

Primarily we were focused on the most active, cooperating co-authors around the selected author (central node). If there is a strong cooperation, then should be a little distance between these two authors. In addition, we were searching for a way how to distinguish more important authors. The idea, among others, comes from scaling circles and important author has around him a circle aura. At least but not last, we are using color edges to highlight relationship between subgroups of authors who cooperate together more often than with the others authors.

When we are drawing the graph of the authors, they are placement on the Golden Spiral. We think it helps people to stay better oriented in the graph and the Golden Ratio composition could be more pleasant for human eyes.

3.1 The Golden Ratio

The Golden Ratio for length to width of rectangles of ≈ 1.618 is considered the most pleasing to the viewers eyes. In another words, if you have a Golden Rectangle and divide the length by the height, you will have the Golden Ratio. Traditionally, mathematicians have denoted the Golden Ratio by the Greek letter phi (ϕ).

Proportions of the Golden Spiral are similar to the Golden Rectangle. The Golden Spiral is a series of quarter-circles drawn in squares. These squares are inscribed in Golden Rectangles. We approximate Golden Spiral by Fibonacci numbers and it becomes Fibonacci Spiral. The Golden Rectangle, Golden Spiral and Fibonacci Spiral are shown in an example below (Fig. 1).

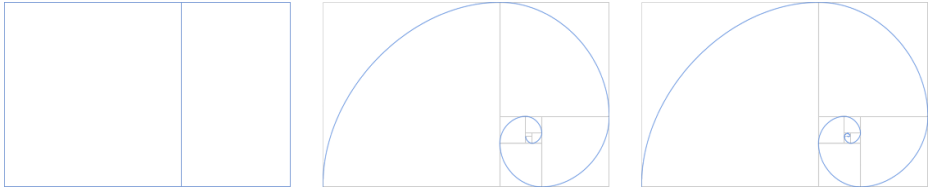


Fig. 1. This figure shows Golden Ratio represented by Golden rectangle, Golden Spiral and Fibonacci Spiral (from left to right)

3.2 The Graphs

The location of the most active co-authors, with the strongest relation with selected author, is close to the center of the Golden Spiral. Their relationships with selected author are the most important. If the relationship of the co-author is less important then his distance from the central node is growing. At least 9 co-authors nodes should be in the graph to realize good looking spiral.

The project Forcoa.NET can identify not only the most important relationships with selected author but also identify somehow important authors in whole graph. These authors has an significant aura around them. Moreover we can identify authors which are working together more often and compartmentalize them into smaller subgroups. These subgroups are highlighted by color. We follow these rules for testing two approaches how to visualize data in the graph. The second approach (Fig. 3) was created as an improvement of the first one (Fig. 2).

4 Experiments

In our experiments we wanted to verify our assumptions by collecting opinions from different people. We decided to do a user testing and also use a questionnaires. When we evaluated results from the user testing and the questionnaires (Table 1), we realized we have couple of things to improve. Our primary focus on a little distance between the most cooperating co-authors with selected author (central node), seems to be right. Without any help or guidelines, the users were choosing as the most cooperating authors the most closest ones to central node. Other assumptions (see Sect. 3) were not convincingly confirmed and the results are shown in Table 1. From the second question an interesting fact was shown. The whole 35.3% of testing users were thinking, the least important co-worker in graph is that one with the least number of edges (Taieb-Maimon).

4.1 First approach

All the nodes in the first approach (Fig. 2) have small aura, which represent their relationship in the graph. If they have a closer relation, not only with a central node but also with the others authors, then their aura is divided into smaller parts by different color. More significant authors have even bigger aura, which can be divided into the same parts as above.

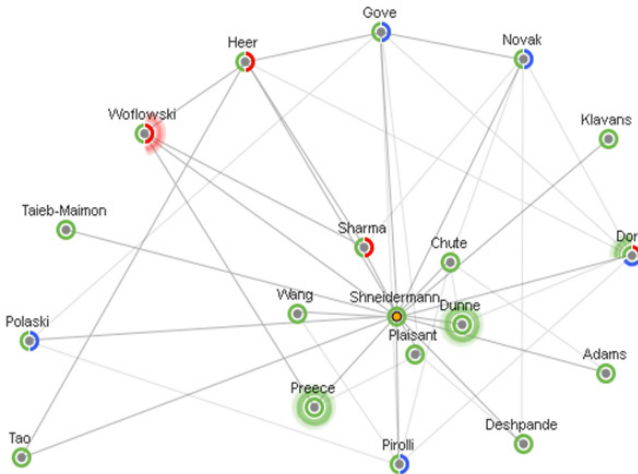


Fig. 2. The Graph - Approach #1

Table 1. The user testing and the questionnaires results from the first approach

Question	Right answers
1. Who are the closest co-workers of Mr. Smith? (<i>Plaisant, Dunne, Chute</i>)	94.1%
2. Who is the least important co-worker of Mr. Smith? (<i>Tao</i>)	58.8%
3. How many subgroups the graph contain? (<i>3 subgroups, green included</i>)	59%
4. Are there people more important than the others? (<i>Dorr, Dunne, Preece, Woflowski</i>)	64.7%
5. Is there someone who is in all subgroups? (<i>Dorr</i>)	35.3%

4.2 Second approach

The second approach (Fig. 3) arise from the first one (Fig. 2), with the aim to enhance results of questions 2 to 5. We decided to transform the visualization of graph even more into flat design. The nodes inner color represents their relation to central

node or another graph. The subgroups are now highlighted by the color of the connecting edges.

The final results were really improved by these changes. Also the second results are more precise because we had twice more testing users and questionnaires answers. Actually we can see decrease by 5% recorded in the first question but we think it could be a result of precision. The second question remained almost without the change. The true is, people were again choosing the node with the least number of edges (Klavans), in 35.1% cases. We consider this answer is not wrong and show us interesting fact. In second and third question we have increase by 20% and in last question the increase is 27%. The second approach is significantly more successful.

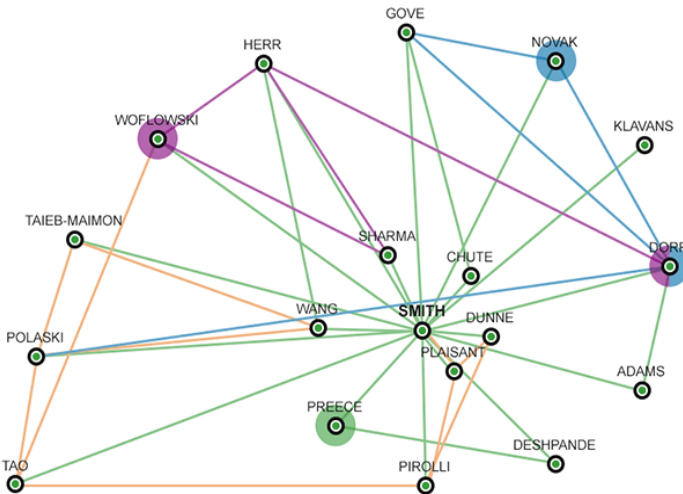


Fig. 3. The Graph - Approach #2

Table 2. The user testing and the questionnaires results from the second approach

Question	Right answers
1. Who are the closest co-workers of Mr. Smith? (Plaisant, Dunne, Chute)	89.2%
2. Who is the least important co-worker of Mr. Smith? (Tao)	59.5%
3. How many subgroups the graph contain? (3 subgroups, green excluded)	81%
4. Are there people more important than the others? (Dorr, Novak, Preece, Wofflowski)	83.8%
5. Is there someone who is more important in more than one subgroup? (Dorr)	62.2%

5 Conclusion

In this article, we introduced our approach how use the Golden Spiral in graph, to visualize small social networks relationships. The main idea about the distance from the center of the spiral and importance of the author, seems to be right. The use of different color for highlights the subgroups and the nodes, brings interesting results. It should be said, during the tests there was not any help or guidelines for our tested users. The results are reached only by an intuition and a common sense.

Future research will focus on more user testing, and verification our ideas with different authors relationships. The challenge will be searching for maximum and minimum quantity of nodes and also their precise placement on the spiral. Moreover we want to implement this Golden Spiral view into our analysis tool Forcoa.NET, to expand the project visualization capabilities.

References

1. M. Kudelka, Z. Horak, V. Snasel, and A. Abraham: “*Social network re-duction based on stability*”. In Computational Aspects of Social Networks (CASoN), 2010 International Conference on, pages 509–514. IEEE, 2010.
2. M. Kudelka, Z. Horak, V. Snasel, and A. Abraham: “*Forcoa.NET: An Interactive Tool for Exploring the Significance of Authorship Networks in DBLP Data*”. In Computational Aspects of Social Networks (CASoN), 2011 International Conference on, pages 261–266. IEEE, 2011.
3. M. Lima: “*Visual complexity: mapping patterns of information*”. New York: Princeton Architectural Press, 2011. Print.
4. E. W. Weisstein: “*Golden Rectangle*”. MathWorld - A Wolfram Web Resource. Retrieved from <http://mathworld.wolfram.com/GoldenRectangle.html>
5. L. C. Freeman: “*Visualizing Social Networks*”. Journal of Social Structure, 1, 2000.
6. C. Correa, C. Tarik, K.-L. Ma, and K. Keeton: “*The Derivatives of Centrality and their applications in Visualizing Social Networks*”. 2009.
7. J. Heer, D. Boyd.: “*Vizster: Visualizing online social networks.*” Proceedings of Symposium on Information Visualization, pages 33–40. Minneapolis, MN: IEEE Press, 2005
8. I. Herman, G. Melançon, M. Scott Marshall: “*Graph visualization and navigation in information visualization: a survey*”. Transactions on Visualization and Computer Graphics, pages 24–43. IEEE, 2000.

Similarity in EEG data using Lempel-Ziv Complexity

Michal Prilepok, Ibrahim Salem Jahan, and Václav Snášel

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
michal.prilepok@vsb.cz, jahan_nw@yahoo.com, vaclav.snasel@vsb.cz

Abstract. Today's big challenge in world is finding efficient technique for recognition between different mental tasks, and distinguish among them. This allow us to use Brain Computer Interface applications to help disabled people to interact with environment and control external devices. In our experiments we used EEG data from National University of Sciences and Technology, Pakistan. We made our experiments on data signals from one subject hand movement task. We applied the faster Fourier Transformer (FFT) to remove the higher frequencies from EEG data, turtle graphics to convert data from measured values into text format for further processing. Then find the similarity between trials by Lempel-Ziv complexity. Our model reached average accuracy up to 52.63% .

1 Introduction

There are several algorithms for similarity analysis and recognition among mental tasks. But still big challenge remains in world to find efficient and fast method for analysis and recognition human mental tasks. We will apply our technique for finding similarity between EEG data and distinguish between two human mental tasks. There was several approaches to classify EEG signal. They are using Support Vector Machines (SVM), L1 regularized logistic regression and non-negative matrix factorization (NMF) [12]. In this paper we will give some glance on EEG, Turtles graphic, Lempel-Ziv Complexity and apply proposed method to find similarity on Electroencephalograph (EEG) data.

2 Introduction to EEG

EEG represents complex irregular signals system that may provide information about underlying neural activities in the brain [2]. The electrical nature of the human nervous system has been recognized for more than a century. It is known that variation of the surface potential distribution on the scalp that reflects functional activities emerging from the underlying brain [7]. This electrical surface potential variation can be recorded by set of electrodes on the scalp, which are measured between pairs of these electrodes and then filterend, amplified, and recorded. The resulting data are called the Electroencephalograph (EEG) [7].

2.1 Source of EEG Generating

The EEG signals are defined as measurements of currents, which are flowing during synaptic excitations of the dendrites of multiple pyramidal neuro cells in the cerebral cortex. When brain cells are activated, the synaptic currents are produced within the dendrites. Normally this current producing a magnetic field, which can be measured by electromyogram (EMG) machines and as electrical field over the scalp measured by EEG systems. The structure of neuro cell in showed on Fig. 1 [11].

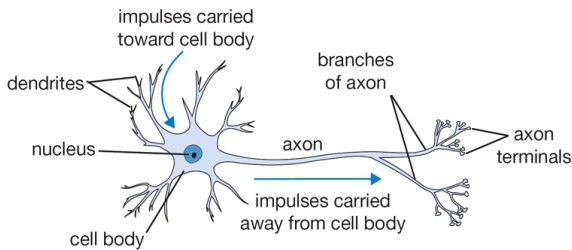


Fig. 1. Structure of Neuro Cell [5]

3 Faster Fourier Transform (FFT)

A Fourier Transform is converting a function from time domain to frequency domain. The faster Fourier Transform (FFT) is algorithm to compute the discrete Fourier transform (DFT) and inverse of DFT. FFT is efficient to compute the discrete Fourier transform (DFT). The FFT is important in field frequency analysis because it takes a discrete signal in time domain and converts this signal to discrete frequency domain representation [3]. FFT and Inverse Fast Fourier Transform (IFFT) algorithms are well know and widely used in several applications for their efficiency [8].

4 Turtle Graphic

The Turtle graphic is an easy way for representation complicated geometric object. Turtle Graphic or L-systems method is used for making the graphics. The basic idea to make graphics by turtle is convert the graphic into sequence of commands, which control turtle and allow to make the specific graphic [9]. Turtle geometry has been used to study and representation many various subjects from simple polygons to complex fractals [4]. To understand turtle graphic geometry, we will explain it on a virtual turtle. The virtual turtle must know own position, facing direction and step size. Each comand can change turtle's position, or heading. For example we have a turtle living on a plane. The turtle position can

be represented by a point P given by a pair of coordinates (p_1, p_2) , similarly its heading can be represented by a vector w given by another pair of coordinates (w_1, w_2) . The step size of the turtle is the length of the vector w . The pair (P, w) is called the turtle's state. Although internally the computer stores the coordinates (p_1, p_2) and (w_1, w_2) [4].

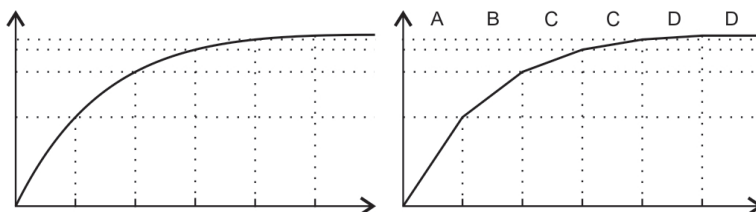


Fig. 2. Data conversion example

On the Fig. 2 we can see a conversion measured raw data into turtle graphic commands. The left figure shows measured raw data in a line chart. The right figure shows interpretation measured values after conversion into turtle graphic commands. The step size is defined by a constant between two measured values we calculate an angle. this angle represents a turtle command. The example commands are on the top of figure. The final commands sequence is ABCDD. In this example we have only four commands. The C and D commands are doubled, because there occurred are equal to previous angle.

5 Lempel–Ziv Complexity

The Lempel–Ziv (LZ) complexity for sequences of finite length was suggested by Lempel and Ziv [6]. It is a non-parametric, simple-to-calculate measure of complexity in a one-dimensional. LZ complexity is related to the number of distinct substrings and the rate of their recurrence along the given sequence [10], with larger values corresponding to more complexity in the data. It has been applied to study the brain function, detect ventricular tachycardia, fibrillation and EEG [13]. It has been applied to extract complexity from mutual information time series of EEGs in order to predict response during isoflurane anaesthesia with artificial neural networks [1].

LZ complexity analysis is based on a coarse-graining of the measurements, so before calculating the complexity measure. The signal must be transformed into a finite symbol sequence. In this study we have used turtle graphic for conversion measured data into finite symbol sequence P . The sequence P is scanned from left to right and the complexity counter is increased by one unit every time a new subsequence of consecutive characters is encountered. The complexity measure can be estimated using the algorithm described in [6] and [1]. We will try describe LZ complexity on a brief example on a binary input. For input 000110100100010 we get following subsequences 0, 001, 10, 100, 1000, 101.

In our experiment we do not deal with measure of the complexity. From the individual subsequences we create a list of them. One list is created for each data file with turtle commands of the compared commands files.

5.1 Comparing data using LZ Complexity and Turtle Graphic Commands

The comparison of the LZ subsequences lists is the main task in our approach. The lists are compared to each other. The main property for comparison is the number of common subsequences in both lists. This number is represented by the sc parameter in the equation 1, which is a metric of similarity between two turtle commands lists.

$$SM = \frac{sc}{\min(c_1, c_2)} \quad (1)$$

Where

- sc is count of common LZ sequences in both dictionaries.
- c_1, c_2 is count of LZ sequences in list of the first or the second file.

The SM value is in the interval between 0 and 1. If $SM = 1$, then thier have most common subsequences in their lists, are equal. The lists have the highest difference, nothing common, when the result value is near $SM = 0$.

6 EEG Experiments

In our experiment we used EEG data available online from National University of Sciences and Technology, Pakistan, We chose Dataset 2 - 2D motion. The EEG raw data was recorded at 500Hz sampling frequency using 19 electrodes. This data set contains several moves trials of hands and legs. In our experiments we used the three trials of left hand back move and imaginary left hand back moves, and two trial of right hand back moves.

6.1 EEG Data Preparation

The EEG data were prepared in following steps. As a first step splited dataset into individual mental tasks, trails and sensors. We got 122 parts. In the second step of our process we applied FFT to transform raw sensor data from time domain into frequency domain. In the frequency domain we removed higher frequencies from data. In the next step we applied Inverse FFT to convert data back from frequency domain into time domain. This filtered data we convert using turtle graphics into text format. For the turtle we used 128 commands in two quadrants - first and fourth. We used only first and fourth quadrant, because the time line goes from left to right and the signal does not go backwards. Each command represents one angle.

After that each EEG trial were prepared by LZ complexity to get LZ subsequences from turtle commands list. For each data trial we created a list of LZ subsequences. We compared training and testing list to find the maximum similarity between EEG trials of same mental task.

7 Experiment Results

We made similarity between the EEG trials for left hand back movement and imaging left hand back movement tasks. Our results are listed in Table 1. Our model reached up average value of similarity 52.36% and accuracy up to 52.63%.

Table 1. Similarity results

	Minimum	Maximum	Average
Correctly identified	30.00%	100.00%	52.63%
Incorrectly identified	0.00%	70.00%	47.37%
True positive rate	0.00%	100.0%	35.53%
False positive rate	0.00%	100.0%	55.26%
Accuracy	30.00%	100.00%	52.63%

On Figure 3 we can see accuracy for sensors. The most accuracy values are between 40% and 60%.

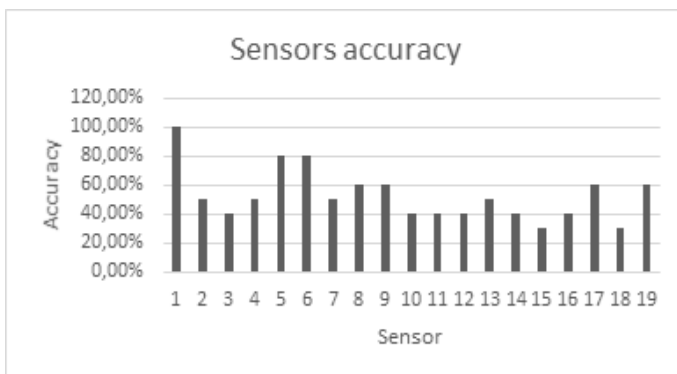


Fig. 3. Sensors accuracy

8 Conclusion and future work

We made our experiments on EEG signals from one subject performing left hand back movement task in three trials, and other trial for imaging left hand back

movement, we applied FFT to EEG data, removing high frequencies, applied inverse FFT, represent EEG data by turtles graphics, then finding the maximum similarity between these trials by LZ compression. Our suggested approach is able to distinguish between to different mental task with average success rate 52.36% and accuracy 52.63%. One of the biggest disadvantages is in average 47.37% incorrectly identified trial.

In the future work we are planning to improve accuracy suggested model with using similarity metric based on data compression for example Normalized Compression Distance and to find better approach.

References

1. D. Abasolo, R. Hornero, C. Gomez, M. Garcia, and M. Lopez. Analysis of EEG background activity in Alzheimer's disease patients with Lempel-Ziv complexity and central tendency measure. *Medical Engineering & Physics*, 28(4):315–322, May 2006.
2. J. Bhattacharya and H. Petsche. Universality in the brain while listening to music. *Proceedings of the Royal Society B: Biological Sciences*, 268(1484):2423–2433, 2001. cited By (since 1996)24.
3. J. Garcia G, C. Castillo L, and C. J. Toledo. Frequency-change analysis of non-linear system using spectrum (fast fourier transform) theory. In *Electrical Engineering, Computing Science and Automatic Control, CCE, 2009 6th International Conference on*, pages 1–6, 2009.
4. R. Goldman, S. Schaefer, and T. Ju. Turtle geometry in computer graphics and computer-aided design. *Computer-Aided Design*, 36(14):1471–1482, 2004.
5. <http://www.wpclipart.com/medical/anatomy/cells/neuron/neuron.png.html>. [Online; accessed 30.7.2013].
6. A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Transactions on Information Theory*, 22(1):75–81, 1976.
7. J. B. Ochoa. Eeg signal classification for brain computer interface applications. Diploma thesis, École Polytechnique Fédérale de Lausanne, March 2002.
8. M. Perko, I. Fajfar, T. Tuma, and J. Puhán. Fast fourier transform computation using a digital cnn simulator. In *Cellular Neural Networks and Their Applications Proceedings, 1998 Fifth IEEE International Workshop on*, pages 230–235, 1998.
9. P. Prusinkiewicz. Graphical applications of l-systems. In *Proceedings on Graphics Interface '86/Vision Interface '86*, pages 247–253, Toronto, Ont., Canada, Canada, 1986. Canadian Information Processing Society.
10. N. Radhakrishnan and B. Gangadhar. Estimating regularity in epileptic seizure time-series data. *Engineering in Medicine and Biology Magazine, IEEE*, 17(3):89–94, 1998.
11. S. Sanei and J. A. Chambers. *EEG Signal Processing*. Wiley-Interscience, Sept. 2007.
12. B. Shin and A. Oh. Bayesian group nonnegative matrix factorization for eeg analysis. *CoRR*, abs/1212.4347, 2012.
13. X.-S. Zhang, R. J. Roy, and E. Jensen. Eeg complexity as a measure of depth of anesthesia for patients. *Biomedical Engineering, IEEE Transactions on*, 48(12):1424–1433, 2001.

Introduction to de novo DNA assembly algorithms

Daniel Robenek, Jan Platoš, and Václav Snášel

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{daniel.robenek.st, jan.platos, vaclav.snasel}@vsb.cz

Abstract. DNA assembly tools are one of the considerably developed and used software. Without efficient algorithms, scientists would not be able to comfortably work with DNA sequences or compare them to find differences. In this paper, most important approaches of DNA assembly software are described and the examples of applicable software packages are shown.

Keywords: DNA assembly, de novo, overlap/layout/consensus, de bruijn graph

1 Introduction

Since the first DNA sequencing method was discovered, reading of its sequences was expensive and long lasting process. The request of scientists to read full human genome has peaked to 1000 genome project. In this project over 4.9GB of data were sequenced using three sequencing technologies [1]. In last ten years, methods of reading DNA sequences has changed. The next-generation methods are cheaper, faster, but brings many difficulties. Unlike Sanger sequencing, next-generation methods gets many of smaller pieces of expected DNA. For example Illumina reads are about 100bp long. And these reads need to be assembled.

Length of DNA may vary from millions (E. coli) to billions (Human) of base pairs. Due to its length, for example human genome has repeats that occur more than 100,000 times each [7]. These repeats are critical parts for DNA assembly, since inappropriate ordering this repeats may devalue assembled sequence.

When the read length is longer than the repeat, it is possible to assembly genome correctly. Next-generation methods only produces very limited read length. Illumina reads are about 200bp long, Roche 454 are up to 500bp and SOLiD produces up to 50bp long reads. For comparison, previously mentioned Sanger sequencing produces up to 1000bp long reads.

To preserve correct order of these repeats, paired-end reads should be used. Paired-end reads may be described as two short reads with known distance between them or single long read with unknown bases in the middle of this read.

Approach to DNA assembly should not only be adapted to DNA length or length of DNA reads. Each sequencing method has different read error rate and even different output format. For example, unlike majority, SOLiD reads are in color space.

In the following chapters, three approaches for the de novo DNA assembly are described with brief introduction of software implementing of them.

2 Greedy

Greedy algorithms are one of the oldest algorithms for DNA assembly. They do not have to explicitly rely on graph, but instead of it, they extend current contig with best match. On each iteration, the best matching read to extend current contig is found. Next, the contig is extended and this process loops until no read is left.

Efficiency of greedy algorithm strongly depends on scoring function that evaluates suitability of usage read for extension. For example this function may be defined as an amount of perfect-matching bases. Finding the best read is the most expensive operation of this algorithm [6].

One of the disadvantages is, that this algorithm can stuck at local maxima. If a read is inappropriately used, a longer contig may break into two shorter contigs. It must also count with false-positive matches, mainly with repeats, which have higher score than the correct one. Of course, there may exist a limit for scoring. If score of any read is below this limit - threshold, the contig is no longer extended. In this case, there exist many contigs instead of one longer scaffold.

First greedy assembler for short reads was SSAKE. It contains lookup table of reads indexed by their prefixes. SSAKE rely on fact, that high coverage of reads should contain at last one error-free read of every part [12], so the assembler look up only perfect matches. It can not operate with paired reads and uniform length of all reads is needed.

3 Overlap/Layout/Consensus

Different approach in comparison with greedy algorithm, overlap/layout/consensus (OLC), uses overlap graph to assembly reads. The flow of the algorithm is divided into three parts.

3.1 Overlap

First part, called overlap, discovers all possible matches of all reads to all reads. For efficiency, the k-mer comparison is used. All k-mers from reads are extracted and the lookup table is created. When there are identical k-mers in the reads, alignment is computed using position of these k-mers. It also exists different algorithms using heuristic or dynamic programming for efficient alignment finding.

This part of algorithm is sensitive to minimal overlap length, k-mer size and the degree of similarity [4]. In case that these parameters are set inappropriately, some overlaps may not be found. For example in case when k is set as too high, there would be many short contigs assembled. This is caused by read errors. To avoid this behavior, k-mers that appear in the DNA only once, or contain error with specified probability, should be fixed. To execute described procedure, it is necessary to find the closest k-mer depending on similarity and fix bases in read.

There also may be some threshold, which specifies acceptable amount of differences in alignment reads caused by read errors. This threshold may vary depending on alignment length [11].

3.2 Layout

In second part of algorithm called layout, the overlap graph is created using previously found followings of reads. Vertices in graph represent fragments and edges represent overlaps. In case of double-strand graph, every fragment represents two vertices joined by edge [11]. The benefit of resulting graph is that it does not have to contain data of reads. Therefore memory consumption may be reduced.

The main task of this graph is to find a maximal set of consistent overlaps.

3.3 Consensus

The last part of algorithm is used to determine the most probable reconstruction of the DNA sequence. The reconstruction must be consistent with layout of fragments.

This is done by multiple sequence alignment. Unfortunately, no effective algorithm to compute multiple sequence alignment is known. Therefore a pair wise alignment is used [2]. Dynamic programming is widely used with this problem [11]. All sequence bases must be loaded in memory in this part of algorithm. Consensus phase is relatively easy to process in parallel.

Celera assembler is an example of software using OLC approach. Originally created for Sanger sequencing method, it was updated to process Roche 454 reads.

Another assembler used, Newbler, developed by 454 Live Science was designed to

process unpaired reads with 100bp length. This assembler uses OLC approach twice. First it generates unitigs from reads. Unitigs are ideally indisputable short contigs created from reads. Second usage of OLC creates sequence from these unitigs [4].

4 De Bruijn graph

The last described approach to DNA assembly is mainly used for short reads, primarily Illumina and SOLiD. It is not based on read comparison as previously described algorithms. It rather extracts all k-mers from reads and works with them. K-mers are used to build directed graph of particular overlaps [13]. This graph is generally called de Bruijn graph.

Basic construction of this graph expect to have one unique k-mer in one vertex. If two reads contains the same k-mer, the vertex count will not change. The only change will be noticed as the change of quantity of this k-mer in the vertex, or edge.

Edges among vertices will appear only if k-mers of this vertices differs only in single base. K-mer of second vertex must be shifted to the left considered to first k-mer of vertex. Since DNA has only 4 bases, each vertex may have 4 input and 4 output edges at the maximum.

Assembly algorithms based on de Bruijn graph suffer from excessive memory requirements especially when source data contains huge amount of read errors [13]. The errors cause many dead ends and bubbles in this graph. These errors may be reduced by using preprocessing of source data [9].

Other approaches to reduce memory complexity are graph error correction and graph reduction. With graph reduction, non-intersecting paths are connected to single vertex. Remove of short tips is also appropriate, since short dead ends are mainly created by read errors. Found bubbles with similar length and bases are reduced to single path, determined by higher coverage of k-mers [13].

Removing paths with low coverage is questionable, because it is not clear if the cause is low read coverage or error-reads. To resolve contigs, it is necessary to find eulerian path in graph [5]. This is a path that traverses each edge exactly once. It is not possible to find single path, due to read errors. DNA contains many repeats that causes crosses in graph. To support sequence reconstruction, paired-reads are used.

There are many advanced assemblers using this approach. Velvet assembler is able to use both the short and the long reads. It uses double-stranded data for graph construction and mate-pairs to find correct eulerian path. Algorithm to reduce errors in the graph is called tour bus. Deeper description of tour-bus can be found in [13]. Velvet is also able to process color-space reads produced by SOLiD sequencer.

The excessive memory requirements of de Bruijn graph solves ABySS assembler by distributed computing. For parallel processing, the reads are divided into k-mers

and these are distributed over available nodes selected by hash function. In the next steps the distributed de Bruijn graph is created. For each node in the graph, the adjacency is computed and a message is sent to neighbors [10]. By this way the connections are known and assembly process may be done.

5 Current Research

In our previous research, we focused on n-gram indexing by tree structures [8]. The novel approaches were introduced and three new improvements of trees are in state of tests. For example, simply by reducing number of unused references in ternary tree structures, we are able to decrease memory requirements up to 40% without any significant slowdown.

Efficient tree structures may be very helpful for DNA assembly software. Since new sequencers also generates quality scores, using tree structures for the range search is appropriate. This is also applicable for sequence alignment software.

Currently, the most used approach is de Bruijn graph. Therefore, we selected widely recognized and accurate assembler, Velvet, as a reference assembler for future comparison and research.

One significant disadvantage of Velvet is, that it is not able to run on highly parallel computer. Using computer cluster, DNA assembly of human genome would be a matter of hours, or minutes instead of days. This approach would also reduce huge memory requirements on single computer.

With this idea, we developed a prototype of first part of Velvet – velveth. This prototype is simpler and more human-friendly for future upgrades. Without any parallelization, it is about 5-8% faster than original velveth. The tests were performed on amount of 120,000 DNA reads of *Staphylococcus aureus*.

The parallelization of velveth was performed in [3], but in comparison of original velveth, it was only 25x faster using 1,800 nodes, with 8 cores each. This speedup is about 0,2% per core. This parallelization is waste of resources.

We have been designing different approach, that should be more efficient. By sophisticated use of divide and conquer approach, the speedup per core should be much higher.

6 Conclusion

The purpose of this paper is to briefly introduce current state of DNA assembly software. Amount, quality and speed of DNA assemblers increase with the requirement to process increasingly larger quantities of sequences. To process these

quantities of data efficiently, highly parallel computing should be used. Next research on this topic should be focused to efficiently parallelize existing accurate de novo DNA assembly algorithms.

Acknowledgement

This work was partially supported by the Grant of SGS No. SP2013/70, VŠB - Technical University of Ostrava, Czech Republic, and by the Bio-Inspired Methods: research, development and knowledge transfer project, reg. no. CZ.1.07/2.3.00/20.0073 funded by Operational Programme Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic.

References

1. Abecasis, G. R., et al. "A map of human genome variation from population-scale sequencing." *Nature* 467.7319 (2010): 1061-1073.
2. Edgar, Robert C., and Serafim Batzoglou. "Multiple sequence alignment." *Current opinion in structural biology* 16.3 (2006): 368-373.
3. Joshi, Nitin, et. al. "Parallelization of Velvet," "a de novo genome sequence assembler"
4. Miller, Jason R., Sergey Koren, and Granger Sutton. "Assembly algorithms for next-generation sequencing data." *Genomics* 95.6 (2010): 315-327.
5. Pevzner, Pavel A., Haixu Tang, and Michael S. Waterman. "An Eulerian path approach to DNA fragment assembly." *Proceedings of the National Academy of Sciences* 98.17 (2001): 9748-9753.
6. Pop, Mihai. "Genome assembly reborn: recent computational challenges." *Briefings in bioinformatics* 10.4 (2009): 354-366.
7. Pop, Mihai, Steven L. Salzberg, and Martin Shumway. "Genome sequence assembly: Algorithms and issues." *Computer* 35.7 (2002): 47-54.
8. Robenek D., Platoš J., Snášel V. "Efficient in-memory data structures for n-grams indexing" *Dateso* (2013): 48-58.
9. Roberts, Michael, et al. "A preprocessor for shotgun assembly of large genomes." *Journal of computational biology* 11.4 (2004): 734-752.
10. Simpson, Jared T., et al. "ABYSS: a parallel assembler for short read sequence data." *Genome research* 19.6 (2009): 1117-1123.
11. Sutton, Granger, and Ian Dew. "Shotgun Fragment Assembly." *Systems Biology: Volume I: Genomics: Volume I: Genomics* 1 (2006): 79.
12. Warren, René L., et al. "Assembling millions of short DNA sequences using SSAKE." *Bioinformatics* 23.4 (2007): 500-501.
13. Zerbino, Daniel R., and Ewan Birney. "Velvet: algorithms for de novo short read assembly using de Bruijn graphs." *Genome research* 18.5 (2008): 821-829.

Stock price prediction based on particle swarm optimization

Lukáš Zaorálek and Tomáš Buriánek

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{lukas.zaoralek, tomas.burianek.st1}@vsb.cz

Abstract. Nowadays, ability to successful predict stock price movement is important area for every participants of stock exchanges such as banks, hedge funds or individual investors. This paper aims to analyze forecast model for stock price prediction using particle swarm optimization (PSO). Included are PSO algorithm, stock price theory with technical analysis and application PSO on a stock market. The paper shows experiments with two selected stocks and several indicators of technical analysis with different settings as inputs for the model and how the forecast model is successful. Finally, it discussed deviations in the markets forecast and theirs impact on the profitability.

Keywords: particle swarm optimization, stock, stock price movement, prediction

1 Introduction

Today, the stock price prediction is one of the most important challenge for both market participants and scientists. The future knowledge of stock price movement can significantly increase investment return on the market and also reduce a risk. [6] An investor can use different approach to reach price prediction based on fundamental or technical analysis. These analysis assist in search an opportunity on the markets and give answer on the question when buy or sell an investment instrument such as the stock. The fundamental analysis include macroeconomic, political and company specific factors. In opposite, the technical analyse uses mathematical and statistical formula such as moving averages, oscillators or correlation.[5] Trading jargon call these mathematical formulas as indicators. These formulas have usually one or more parameters that significantly influence a result. Market participants are using technical analyse to recognize graphics patterns in a charts of indicators. Based on patterns traders make decision to trade a stock. [1] In this paper we focused on technical analyse and optimize parameters of several indicators in a way to predict stock price. First section introduces stock markets and technical analyse. Next, particle swarm optimization is reviewed. Third and last chapter discussed about the experiment and results.

2 Stocks markets

The stock market also known as the equity market is a place (physical or electronic) where stocks, and derivatives in listed companies are traded. This place also known as stock exchange. The one of the biggest stock exchanges is NYSE (New York Stock Exchange). There are listed more than 2000 companies. The following statement describes basic concept how stocks are traded on the exchange and then technical analysis is discussed in more details.

2.1 Continuous double auction

The stock market is based on a continuous double auction (CDA). The double auction consists of buyers and sellers which submit their trade orders to the market. Buyers enter their bids and sellers enter their offers simultaneously as limit order. Trades are filled (or executed) if the highest bid exceeds or is equal to the highest offer. Traders who wanted to buy or sell immediately can submit market order. Limit order is stored in an order book if these limit order is not filled immediately.

2.2 Technical analysis

Technical indicator is statistical information to determine future trends of stock price. It is mathematical calculation from previous historical data of price and volume. [4] Obtained values such as high, low, close and open price of a day are used to predict future price change, movement or direction. Many technical indicators were invented till now. It can be divided to three subgroups: Volume indicators, Oscillators, Trend indicators. These indicators are used by active traders in the market. They are using various combinations of indicators to make decisions. The main task for us is how to combine the set of picked indicators together, to acquire acceptable price prediction. The following table 1 introduces several indicators and their formula.[7]

3 Particle Swarm Optimization

Particle Swarm Optimization (PSO) is one of the evolutionary computation techniques. It was introduced by Russ Eberhart and James Kennedy in 1995. This algorithm is based on social behaviors inspired by bird flocks or other swarm animals with collective behavior. [3] The main goal is to find the optimal solution over the search space for defined fitness function of solving problem. Model of PSO is based on swarm. Swarm is a set of particles, where each particle is defined by its position and velocity. All particles fly through the search space and try to find optimal position for our problem with best fitness value. The movement of each particle is influenced by its best position (local optimum) and by the best position of all particles (global optimum). The influence is realized as change of position vector of each particle. [2]

Table 1. Indicators

Indicator name	Formula	Description
Simple moving average	$SMA = \frac{p_m - p_{m-1} \dots p_{m-(n-1)}}{n}$	p is close price of current day
Exponential moving average	$EMA = (p_t * \alpha) + EMA(p_y) * (1 - \alpha)$	p_t is close price of current day and p_y is close price previously day
Relative strength index	$RSI = \frac{EMA(p_t - p_y)}{EMA(p_y - p_t)}$	p_t is close price of current day and p_y is close price previously day
Momentum	$MOM = p_t - p_n$	p_t is close price of current day and p_n is close price n-days ago
Commodity channel index	$CCI = \frac{1}{0.015} \frac{T_p - SMA(T_p)}{\sigma(T_p)}$, $T_p = \frac{p_h + p_l + p_c}{3}$	p_h is higher price of current day, p_l is lower price of current day, p_c is close price of current day, $\sigma(T_p)$ is mean absolute deviation
Money flow index	$MFI = \frac{PMF}{PMF - NMF} * 100$, $PMF = T_p * v$, $NMF = T_p * v$, $T = \frac{p_h + p_l + p_c}{3}$	p_h is higher price of current day, p_l is lower price of current day, p_c is close price of current day

$$v_{ij}(t + 1) = \omega v_{ij}(t) + c_1 r_{1j}(p_{ij} - x_{ij}(t)) + c_2 r_{2j}(p_{gj} - x_{ij}(t)) \tag{1}$$

$$x_{ij}(t + 1) = x_{ij}(t) + v_{ij}(t + 1) \tag{2}$$

For each particle i on each dimension j we obtain new velocity from previous velocity and previous position of particle, local optimum p_{ij} and global optimum p_{gj} , Parameters c_1 and c_2 are acceleration constants, w is inertia weight and r_1 and r_2 are the uniformly generated random numbers in the range of $[0, 1]$. New position of particle is obtained from previous position vector and new velocity.

In every evolution step we attract every particle on every dimension to its local optimum and global optimum. In each step we obtain fitness value from given fitness function for each particle and we determine if there are better local optimum points and some global optimum point. The algorithm is described as follows:

1. Initialization of M particles:
 - (a) Random initialization of positions and velocities
 - (b) Evaluation of fitness value
 - (c) Finding best local position for each particle and best global position
2. Evolutionary step for M particles:
 - (a) Obtaining new velocity from equation (1)
 - (b) Obtaining new position from equation (2)
 - (c) Evaluation of fitness value
 - (d) Finding best local position for each particle and best global position
 - (e) Checking end condition of desired minimum error or maximum of evolutionary steps not satisfying, repeat from step 2
3. Return best global position as best solution

4 Experiments

In this section is introduced data preparation and algorithm of the approach.

4.1 Data preparation

We have picked indicators: RSI, MOM, MFI, EMA with period 30 and 15, SMA with period 30, 15 and 10, MAMA with period 30 and 15 to our experiments, described in table 1. The data for stock prediction has been collected for Yahoo (stock symbol YHOO) and Microsoft (stock symbol MSFT) from 12.04.1996 to 12.07.2013 (each price per day). We have prepared experimental data from picked technical indicators and open, high, low, close prices and stock volumes. The data were divided to two subsets 70% of data for training set and 30% of data for testing set. Normalization for data is carried out to range between -1 and 1.

4.2 Algorithm

Technical indicators are combined by linear combiner defined in equation 3.

$$p(m) = \sum_i (w_i * in_i) \quad (3)$$

Each indicator in_i is used as input to linear combiner $p(m)$, multiplied by each weight w_i parameter and summed. Output of this is for us predicted stock price. Our objective is to optimize these weights for each indicator. Optimization is made using Particle Swarm Optimization by minimizing mean square error (MSE). MSE is used as fitness function, to evaluate each particles error as a measure of optimal solution. Algorithm is described below:

1. Initialize PSO, random particles weights
2. Extracting M input patterns from N input indicators from past stock data.
3. PSO optimization of linear combiner.
 - (a) For each particle:
 - i. For each m pattern from input, predicted output $p(m)$ is computed, where weights are from current particle.
 - ii. Error $e(m)$ of predicted output $p(m)$ compared to desired output $d(m)$ is obtained $e(m) = p(m) - d(m)$
 - iii. Particle error from MSE is obtained from M errors
 - (b) In each evolution step, weights are optimized by updating velocity and position of particles
 - (c) Optimization is stopped after finding minimum error of global best position.
4. Best global position from PSO is desired as best solution of weights for linear combiner.

5 Results

The table 2 shows best fitness and mean absolute percentage error (MAPE) per stock per day to prediction. The best prediction day for both of stocks is first day after actual day. Best fitness decrease with number of days to predict and also confirm that first day after actual day is best for prediction. The MAPE shows the MSFT is better stock to prediction than YHOO. The figure 1 and 2 illustrate actual and prediction stock price.

Table 2. Comparision of MSFT and YHOO stocks prediction with best fitness and MAPE

Stock symbol	n-th day to prediction	Best fitness	MAPE
MSFT	1	0.002129	2.3799%
MSFT	3	0.0046125	3.047%
MSFT	5	0.007077	3.6023%
MSFT	7	0.009574	4.0491%
YHOO	1	0.002774	7.2048%
YHOO	3	0.004932	10.0904%
YHOO	5	0.006789	12.5288%
YHOO	7	0.008347	14.6139%

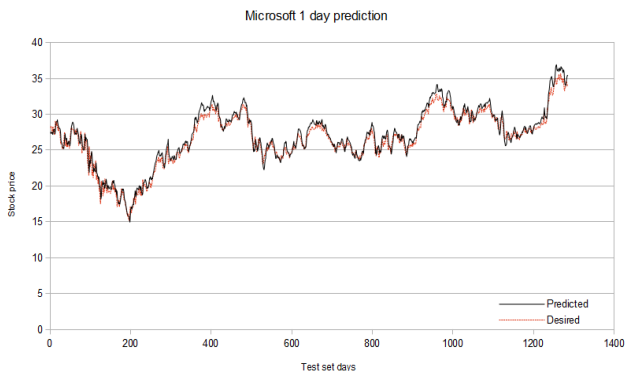


Fig. 1. MSFT 1st day

6 Conclusion

In this paper we show stock price prediction. We use several indicators of technical analyse and particle swarm optimilization to predict price. Partical swarm

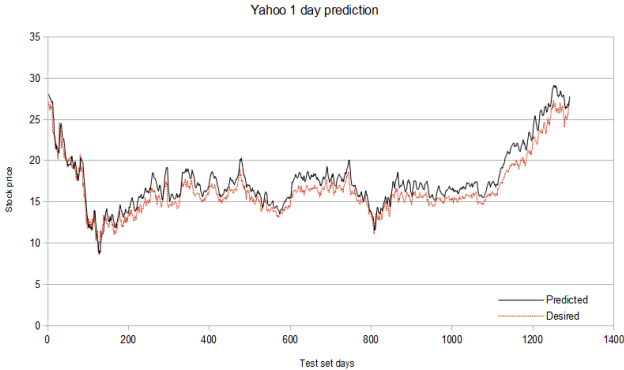


Fig. 2. YHOO 1st day

optimilization found optimal weights of each indicators in linear combiner to calculate future stock price. By testing on test dataset, we have verified efectivity of our forecast model. We have acquired very interesting and sucesful predictions, especially for Microsoft stock market with one day prediction.

References

1. Franklin Allen and Risto Karjalainen. Using genetic algorithms to find technical trading rules. *Journal of Financial Economics*, 51(2):245 – 271, 1999.
2. R. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In *Micro Machine and Human Science, 1995. MHS '95., Proceedings of the Sixth International Symposium on*, pages 39–43, 1995.
3. R.C. Eberhart and Yuhui Shi. Particle swarm optimization: developments, applications and resources. In *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, volume 1, pages 81–86 vol. 1, 2001.
4. Fernando Fernndez-Rodrguez, Christian Gonzlez-Martel, and Simn Sosvilla-Rivero. On the profitability of technical trading rules based on artificial neural networks:: Evidence from the madrid stock market. *Economics Letters*, 69(1):89 – 94, 2000.
5. Jovita Nenortaite and Rimvydas Simutis. Stocks trading system based on the particle swarm optimization algorithm. In Marian Bubak, GeertDick Albada, PeterM.A. Sloot, and Jack Dongarra, editors, *Computational Science - ICCS 2004*, volume 3039 of *Lecture Notes in Computer Science*, pages 843–850. Springer Berlin Heidelberg, 2004.
6. E. Nourani, A.-M. Rahmani, and A.H. Navin. Forecasting stock prices using a hybrid artificial bee colony based neural network. In *Innovation Management and Technology Research (ICIMTR), 2012 International Conference on*, pages 486–490, 2012.
7. Yudong Zhang and Lenan Wu. Stock market prediction of s&p 500 via combination of improved bco approach and bp neural network. *Expert Systems with Applications*, 36(5):8849 – 8854, 2009.



euROPEAN
social fund in the
czech republic



EUROPEAN UNION



MINISTRY OF EDUCATION,
YOUTH AND SPORTS



OP Education
for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT

SoftComp Workshop

SoftComp workshop – organized as a part of WOFEX 2013 workshop – have been realized as a part of SoftComp project (CZ.1.07/2.3.00/20.0072). This project is co-financed by ESF and Czech state budget.



Acknowledgement

Following papers were supported by the Development of human resources in research and development of latest soft computing methods and their application in practice, reg. no. CZ.1.07/2.3.00/20.0072, co-financed by ESF and state budget of the Czech Republic:

- Alisa Babskova. *Finding Evolution and Measuring of Stability of Co-Authors Communities Formed by Terms on DBLP*
- Patrik Dubec and Jan Plucar. *Comparison of graph compression algorithms*
- Michal Holis, Lenka Skanderova, Martin Placek, Jiri Dvorsky, and Ivan Zelinka. *Solving Steel Alloying Using Differential Evolution and SOMA*
- Lukas Rapant. *Comparison of Bio-inspired and Mathematical Optimization Methods on a Graph Drawing Problem*
- Daniel Stříbný and Zbyněk Složil. *Architectures and Features of Modern Honeypots*
- Lukáš Vojáček. *Growing Neural Gas – A Parallel Approach*

Finding Evolution and Measuring of Stability of Co-Authors Communities Formed by Terms on DBLP

Alisa Babskova*

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
alisa.babskova.st@vsb.cz

Abstract. Finding and evaluation of communities in social or collaboration networks is still a big challenge. The purpose of this paper is to find the method for evaluation of the stability of authors interests in the communities extracted from DBLP. Find the method for identification of dynamic or static communities in the time. The experiments were demonstrated on the network of co-authors.

Keywords: Dynamic Social Network, DBLP, co-authors, evolution

1 Introduction

The study of the dynamic evolution is relatively new subject in the research of the social communities. The research of this paper is focused to study the communities extracted from the DBLP dataset and their dynamic grow in time.

The aim of this paper was to develop a methodology for finding, tracking, analysing and evaluating the development of the groups of authors who deal with the areas specified by chosen terms. We can see whether this area is still developing, expires, is stable or promising. The results of this paper could be used by researchers to point their professional interest.

2 Social Networks and Digital Bibliography Library Project

A social network (SN) is a set of people or groups of people with similar pattern of contacts or interactions such as friendship, co-working, or information exchange [3]. The World Wide Web, citation networks, human activity on the internet,

* This paper is based on article named ‘Evolution of Co-Authors Communities Formed by Terms on DBLP’. Therefore I would like to thank co-authors, namely Pavla Dráždilová, Jan Martinovič, Václav Svatoň, Václav Snášel for cooperation with this article and agreement to its publication.

physical and biochemical networks are some examples of social networks. Social networks are usually represented by graphs, where nodes represent individuals or groups and lines represent relations among them.

Social network analysis (SNA) is a collection of methods, techniques and tools that aim to analyse the social structures and relational aspects of these structures in a social network [2]. The study of social networks is a quite old discipline. Many studies oriented to the analysis of social networks have been provided. The datasets used in these studies are obtained by using questionnaires.

DBLP (Digital Bibliography Library Project) is a computer science bibliography database hosted at University of Trier, in Germany. It was started at the end of 1993 and listed more than 2.1 million publications in January 2013. DBLP has been a credible resource for finding publications, its dataset has been widely investigated in a number of studies related to data mining and social networks to solve different tasks such as recommender systems, experts finding, name ambiguity, etc. Even though, DBLP dataset provides abundant information about author relationships, conferences, and scientific communities, it has a major limitation that is its records provide only the paper title without the abstract and index terms.

3 Dynamic network analysis

Dynamic network analysis (DNA) varies from traditional social network analysis. DNA could be used for analysis of the non static information of nodes and edges of social network. DNA is a theory in which relations and strength of relations are dynamic in time and the change in the one part of the system is propagated through the whole system, and so on. DNA opens many possibilities to analyse and study the different parts of the social networks. We can study behaviour of individual communities, persons or the whole graph of the social network. The paper is focused to analyse the behaviour of communities extracted from DBLP and divided by time. The proposed approach which use dynamic metrics is inspired by work of Palla et al. [1].

The auto-correlation function $C(t)$ is used to quantify the relative overlap between two states of the same community $A(t)$ at t time steps apart:

$$C(t) = \frac{|A(t_0) \cap A(t_0 + t)|}{|A(t_0) \cup A(t_0 + t)|}, \quad (1)$$

where $|A(t_0) \cap A(t_0 + t)|$ is the number of common nodes (members) in $A(t_0)$ and $A(t_0 + t)$, and $|A(t_0) \cup A(t_0 + t)|$ is the number of nodes in the union of $A(t_0)$ and $A(t_0 + t)$.

The stationarity of community is defined as the average correlation between subsequent states:

$$\zeta = \frac{\sum_{t=t_0}^{t_{max}-1} C(t, t+1)}{t_{max} - t_0}, \quad (2)$$

where t_0 denotes the birth of the community, and t_{max} is the last step before the extinction of the community. Thus, $(1 - \zeta)$ represents the average ratio of members changed in one step.

Authors of the paper [1] found that the auto-correlation function decays faster for the larger communities, showing that the membership of the larger communities is changing at a higher rate. In contrast, they said that small communities change at a smaller rate with their composition being more or less static. The stationarity was used to quantify static aspect of community evolution.

4 Evolution of Co-authors Communities

To create our experiments and to count dynamic metrics we generate DBLP subgraphs of selected terms for each year in which this term occurs. This final set of subgraphs is input for our experiments and to count dynamic metrics.

4.1 Extraction of Communities of Co-authors in Time

Extraction of Communities of Co-authors in Time is shown in paper [4]. As a result we had for every year $y \in Y$ we create graph $G_y(Authors_y, E)$, where E represents strength of authorship. Dynamic metrics described in the Section 3 are generally metrics used to evaluate the characteristics of the community. About such community, we have to know that it changes over time and also we should have information on how the community looked at each time step of its existence. Therefore to get the information about the communities and their changes in time from subgraph of the authors, we need to execute a series of steps which are described below.

4.2 Algorithm for Finding Component Evolution in Time

- (I) *Creating the longest continuous consecutive time chain of graphs G_y*
 Input graphs may have different time intervals between them. But for the next step we need to choose the longest consecutive time period with one year interval.
 For example:
 Input graphs: $G_{1998}, G_{1999}, G_{2002}, G_{2003}, G_{2004}, G_{2005}, G_{2006}, G_{2007}, G_{2012}$.
 For processing we use this set of graphs: $G_{2002}, G_{2003}, G_{2004}, G_{2005}, G_{2006}, G_{2007}$.
- (II) *Finding connected components of the subgraph*
 Graphs from the previous step are non connected. We search for all the connected components to get components for each year with which we will continue to work.
- (III) *Create chain of the connected components across all time steps*
 1. We choose the first largest component c from the graph in the first time step.

2. According to the following rules we select next component (follower) in the next time step based on the current component c . We denote this component as similar component. We are looking for the components which has the biggest number of the same nodes as the current component c and for selection we have to choose one of the following options:
 - (a) If only one similar component is found we denote it as follower.
 - (b) If more than one similar components are found we denote the biggest one as follower.
 - (c) If no component is found we choose as a follower the biggest existing component in this time step.
3. Step 2 is repeated for each time step except the last one.

Basically we are talking about the components that consist of the DBLP authors and links between them which are formed on the basis of the common interest - the same terms in the titles of their articles. Therefore we can say that our components are the communities of the co-authors. Due to the above described algorithm, we prepare the set of consecutive components. We assume that this set represents the development of one community over time.

This idea allows us to calculate dynamic metrics described in the Section 3. Recall that the auto-correlation is calculated for each of the two states of the same community, followed with computed value of stationarity.

5 Experiments

To demonstrate experiments, we choose terms: "social network", "dynamic social network", "social network analysis". Basic properties of the communities found for each set are described in the table, where we present the count of time steps for each community.

Table 1. Communities of co-authors developed in time

Terms	Count of time steps	Year from	Year to
social network	17	1997	2013
dynamic social network	10	2003	2013
social network analysis	17	1997	2012

Evolution of communities of co-authors in the time are demonstrated on the left side of Figure 1. These figures show changes of counts of members of each community in time.

On the right side of Figure 1 we can see a development of the three communities, which published in similar areas, namely "social network", "dynamic social network" and "social network analysis". If we look at the change of the curves of authors in communities that deal with "social network" and "social network analysis", we will notice that curves from 1997 to 2009 look similar. In 2009,

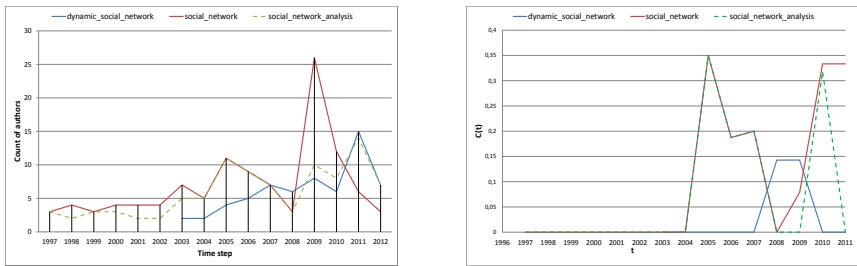


Fig. 1. Evolution and Auto-correlations of communities of co-authors for the terms "social network", "dynamic social network", "social network analysis"

we can notice a great interest in the generic term "social network". According to information shared by Facebook provider in 2009¹, there was the largest detected increase of new users on Facebook. In 2009, around 150 million new users have joined the social networking site Facebook. In the following years, the number of newly connected users varied from 5 to 50 millions per year.

Since 2009, interest in generic term "social network" began to decline strongly. On the other hand, interest in terms "dynamic social network" and "social network analysis" had increased. At the same time, these two curves began to grow similarly.

On the left side of Figure 1, we present auto-correlation values for communities "social network", "dynamic social network" and "social network analysis". The higher the auto-correlation is, the more authors in the community in these time periods had stable interest in publishing together with someone else. This means in our case publishing together in the same area of interest that was initialized by the terms. According to the auto-correlation curves, there was stable interest in "social network" in 2004 which then continuously decreased until 2009. From 2009 onwards we can see a stable growth of interest in publishing in "social network". From 2007 to 2010, there is evident growth of interest in the field of "dynamic social network". However, it is smaller than that of the generic term "social network". For the community "social network analysis", we can follow a similar stability evolution of the authors who published in the area of "social networks".

In the Table 2, we show the values of stationarity for all communities, which we analysed in our experiments. In general, this value characterizes the degree of variability of community in time. The larger the value of stationarity is, the more the community is stable and static. On the other hand, the smaller value indicates a community more dynamic and more changeable in time.

¹ Number of active users at Facebook over the years, <http://news.yahoo.com/number-active-users-facebook-over-230449748.html>

Table 2. Stationarity of communities of co-authors for terms "social network", "dynamic social network", "social network analysis"

Communities of co-authors for the terms	Stationarity
social network	0,0926946271929825
dynamic social network	0,0317460317460317
social network analysis	0,0659801136363636

6 Conclusion

In the paper, the analysis of evolution of co-authors in the communities was presented, with the focus to their growth. The method for identification of dynamic or static communities in the time was presented. Experiments have been demonstrated on the network of co-authors of DBLP. Naturally, presented methods can be used for other different networks and another types of communities.

The step Number III is one of the most important steps in the algorithm presented in the Section 4.1, because it defines which components represent an image of one component in different time periods. In future, we want to enrich our experiments by changing this step of the presented algorithm. Together with condition for a particular user incorporated into this step, it gives a completely different view on the issue of selecting the components. Analysis of the evolution of community formed around a user brings the opportunity to research and analyse not only dynamic properties of the community itself but also the possibility of studying the characteristics of the users or the analysis of evolution in individual cases.

7 List of publications

Evolution of Co-Authors Communities Formed by Terms on DBLP

A. Babskova, P. Dráždilová, J. Martinovič, V. Svatoň, V. Snášel *DATESO*, 2013

References

1. G. Palla, Albert-lászló Barabási and Tamás Vicsek, and B. Hungary, Guy I. Quantifying social group evolution. *Nature*, (46):664667, 2007.
2. J. Scott. *Social Network Analysis*. Newbury Park CA: Sage, 1992.
3. F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks, Feb 2004.
4. A. Babskova, P. Dráždilová, J. Martinovič, V. Svatoň, V. Snášel. Evolution of Co-Authors Communities Formed by Terms on DBLP, *DATESO2013*, 2013

Comparison of graph compression algorithms

Patrik Dubec and Jan Plucar

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{patrik.dubec, jan.plucar}@vsb.cz

Abstract. In our late research we have presented several papers concerning the issue of drawing large scale graphs. This issue arose when dealing with the problem of social network analysis. However, during this research we have been faced with yet another problem. Large scale graphs have too many vertices and edges to be stored in the memory. Graph could be stored partially in the memory and partially on the hard drive. Such approach considerably slows down visualization of a graph due to the high number of search operations between main memory and the hard drive. This paper tries to list and compare approaches to graph compression algorithms that could ensure good enough compression of a graph in order to work with this graph in the main memory.

1 Introduction

There are many research projects and applications that are used for social network analysis, one of them being solved under the FP7 Indect project. The task of the project is to crawl social networks, download content of the communication among users and analyze this communication. During this analysis the software tries to identify entities and relations between them. Information about entities is stored in a form of graph. Graph is an ordered pair $G = (V, E)$ composed of a set of vertices and/or edges, where V represents set of vertices $V = \{1, \dots, n\}$ and E represents set of edges $E = \{e_1, \dots, e_m\}$. Pair of vertices is linked together using edges. These vertices are called adjacent (neighbors). Vertices represent objects of a real world and edges represent relations between them.

Once the analysis is complete and information is stored in graph, the application tries to visualize the graph. The application uses several graph layout placement algorithms depending on circumstances. The key problem here is the number of queries when trying to obtain adjacency list for each vertex. Large scale graphs are too large to be stored in the main memory and thus we come to a performance problem when querying graphs stored on the hard drive.

2 State of the art

In past years many scientist focused on efficient graph compression algorithms. However, there is one research presented by Boldi and Vigna (BV) [7] in 2004 that is considered to be very efficient. BV results are frequently used as a metric for evaluation of algorithm performance. Their method is likely to achieve around 3 bits per edge (may vary from dataset to dataset), at link access time below 1ms at their 2.4GHz Pentium4 machine. Studying several datasets, Boldi and Vigna noticed high level of similarity. Two adjacent vertices tend to have similar list of relations (edges). Similarity is much more concentrated than it was previously thought. Two lists have either almost nothing in common or they share large number of same relations. To exploit redundancy between two lists, BV introduces “modified” version of one of the lists called *reference list*. Reference list is used to denote which of its relations are copied to current list and which are not.

Table 1. Graph representation using adjacency list

Node	Outdegree	Adjacent vertices
15	11	13, 15, 16, 17, 18, 19, 23, 24, 203, 315
16	10	15, 16, 17, 22, 23, 24, 315, 316, 317
...

Table 2. Graph representation using reference list and copy list

Node	Outdegree	Ref.	Copy list	Extra nodes
15	11	...		13, 15, 16, 17, 18, 19, 23, 24, 203, 315
16	10	15	0111001101	22, 316, 317
...

Copy lists contain sequences of 0s and 1s. These sequences are encoded using run-length encoding. Relations that did not occur on copy list (Table 2 – extra nodes) are also stored. Compression algorithm allows selecting different references lists to find the most suitable solution.

Grabowski and Bieniecki [2] present “similarity of successive lists” (SSL) algorithm similar to BV that works in two phases: preprocessing and final compression. Grabovski and Bieniecki introduce copy lists that are no longer binary, but use several flags instead. These flags depict similarity of current list to reference list. Another change to BV is that reference list is always previous adjacency list. SSL demonstrates very good compression ratios. However, there is a tradeoff in a form of slow searching such a compressed structure. It is in fact only slightly better than reading data from hard drive.

Very good compression results were achieved in the work of Asano et al. [3]. Asano et al. employ binary adjacency matrix where 1 stands for an edge between two vertices. Next step is to identify one of the six types of blocks in the matrix: singleton,

horizontal, vertical, L-shaped, rectangular and diagonal blocks. For demonstration see Fig. 1.

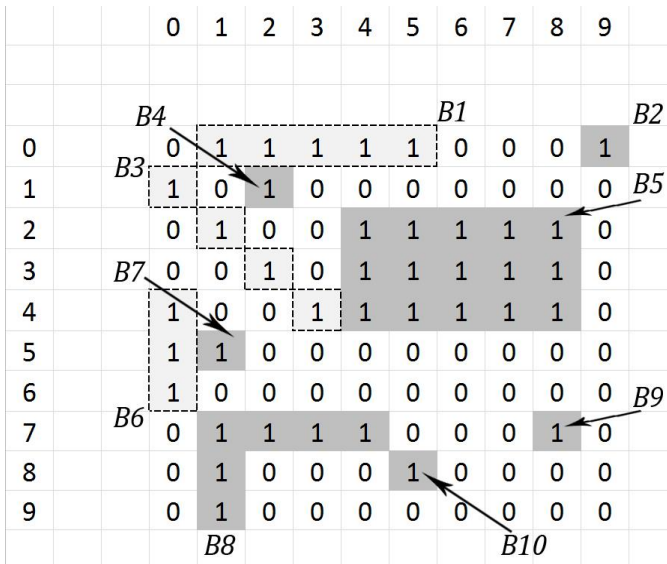


Fig. 1. Adjacency matrix, 10 blocks of different types have been found

Each block is represented by its first element, its type and size. The algorithm compresses the data of these blocks separately using either Elias coding [4] or one of the codes from Boldi and Vigna code family [5]. On average, size of data compressed by Asano et.al. is reduced to 70% compared with Boldi and Vigna on the same data sets. Good compression ratio is however drawn back with the retrieval speed for adjacency list. Worst results were obtained in “eu-2005” dataset (explained in section 3) with retrieval times over 28ms. The application was written using Java programming language and it was tested on a computer with Core2 Duo E6600 (2.40GHz) processor. Authors do believe that retrieval times could be significantly improved and will try to do so in the future work.

Excellent compression results were achieved by Buehrer and Chellapilla (BC) [6], who used grammar-based compression. They search for bipartite graphs into witch they introduced new virtual node to mitigate unnecessary edges, see Fig. 2.

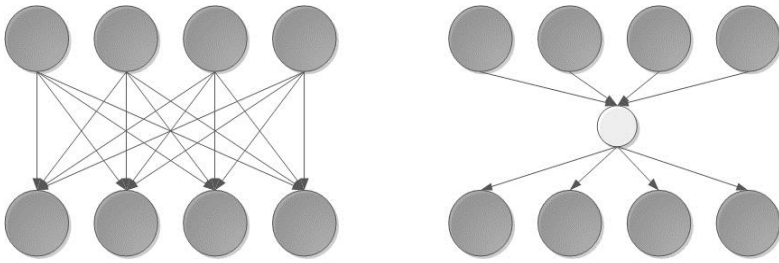


Fig. 2. Original graph on the left was transformed to the graph with a virtual node

Unfortunately no access times were published in BC paper [6], but according to the more recent findings in [7], access times should be better than in BV research.

Apostolico and Drovandi (AD) [8] proposed an alternative graph ordering method. It performs breadth-first traversal of graph and index each node according to the order in which it is expanded. AD obtained very well compressed graphs, usually smaller by 20–30% than those from BV. They also introduced a fast query that is able to check existence of edge between two vertices without the need to generate entire adjacency list.

3 Results

In order to compare results we need to run all mentioned algorithms on the same dataset. Fortunately there are several public datasets that are commonly used for this comparison published on <http://law.di.unimi.it/datasets.php>. Basic information about datasets can be seen in Table 3. For more information about the datasets visit the link mentioned above.

Table 3. Datasets commonly used for testing the performance of compression algorithms (<http://law.di.unimi.it/datasets.php>).

Dataset	Vertices	Edges	Avg Degree
crn-2000	325 557	3 216 152	9.88
eu-2005	862 664	19 235 140	22.30
in-2004	7 414 866	16 917 053	12.23
uk-2002	18 520 487	298 113 762	16.10
arabic-2005	22 744 080	639 999 458	28.14

Compared sizes in bits per link are shown in Table 4. Parameter R in BV stands for maximum reference count (for full description of this parameter see [1], however, a small value for R is likely to produce worse compression, but shorter retrieval times). All the algorithms performed as expected and managed to compress given graph structure. There were two solutions that exceeded others in the matter of size in bits per link. One of the solutions comes from Asano et. al. This solution demonstrates overall

very good results compressing data mostly under 2 bits per link. Second solution is Grabowski and Bieniecki algorithm that demonstrates even better compression.

Table 4. Compared sizes in bits per link BV- Boldi and Vigna, GB – Grabowski and Bieniecki, Asano, AD – Apostolico and Drovandi, BC – Buehrer and Chellapilla

Dataset	BV(R=3)	BV(R= ∞)	GB (SSL)	Asano	BFS	BC
crn-2000	3,56	2,84	-	1,99	2,64	-
in-2004	2,82	2,17	1,10	1,71	2,19	-
eu-2005	5,17	4,38	2,28	2,78	3,48	2,90
uk-2002	3,00	2,22	-	-	2,62	1,95
arabic-2005	2,81	1,99	-	-	2,30	1,81

The size of the compressed graph is not the only measurement that is important. Another view of the compression problem is the retrieval time of adjacency list in different algorithms, see Table 5. As we can see, algorithms that previously exceeded when measuring compression size are very time consuming. The worst results were obtained using Asano et. al algorithm with retrieval time nearing 12ms.

Table 5. Retrieval time of adjacency list

Dataset	BV	GB (SSL)	BFS	Asano
eu-2005	0,24 μ s	18,59 μ s	0,219 μ s	12 ms
in-2004	0,21 μ s	20,77 μ s	0,199 μ s	0,79 ms
uk-2002	0,22 μ s	-	0,264 μ s	-
arabic-2005	0,193 μ s	-	0,168 μ s	-
crn-2000	0,136 μ s	-	0,086 μ s	0,62 ms

3 Conclusion

The purpose of this paper was to find a graph compression algorithm that would help us enhance performance in our research in the field of social network analysis. We have found and compared several graph compression algorithms. In a matter of compression and time performance, we do believe that BFS algorithm published Apostolico and Drovandi would be the most suitable solution. During the test of this algorithm we were able to obtain good enough compression rate – mostly under 3bits per link and fast adjacency list retrieval time. We will also follow research of Grabowski and Bieniecki whose compression algorithms have shown good result in compression rates. The drawback of their approach is a bit higher adjacency list retrieval time. This problem can be solved by the introduction of another algorithm that is much faster, but with slightly worse compression.

References

1. Boldi, P., Vigna, S.: The webgraph framework I: compression techniques. WWW '04 Proceedings of the 13th international conference on World Wide Web (2004), 592-602
2. Grabowski, S., Bieniecki, W.: Tight and simple Web graph compression.; CoRR(2010)
3. Asano, Y., Miyawaki, Y., Nishizeki, T.: Efficient compression of web graphs, 14th Annual International Conference, COCOON (2008), 1-11
4. Elias, P.: Universal Codeword Sets and Representations of the Integers, IEEE Transactions on Information Theory (1975), 194-203
5. Boldi, P., Vigna, S.: Codes for the World Wide Web, Internet Math (2005), 407-429
6. Buehrer, G., Chellapilla, K.: A scalable pattern mining approach to web graph compression with communities, International Conference on Web Search and Data Mining (2008), 95-106
7. Karande, C., Chellapilla, K., Andersen, R.; Speeding up algorithms on compressed web graphs, ACM International Conference on Web Search and Data Mining (2009), 272-281
8. Apostolico, A., Drovandi, G.; Graph compression by BFS, Algorithms (2009), 1031-1044

Solving Steel Alloying Using Differential Evolution and SOMA

Michal Holis, Lenka Skanderova, Martin Placek, Jiri Dvorsky, and Ivan Zelinka

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{michal.holis, lenka.skanderova, martin.placek,
jiri.dvorsky, ivan.zelinka}@vsb.cz

Abstract. This paper proposes method for solving steel alloying problem using evolution algorithms SOMA and differential evolution. Both algorithms belong to the family of the evolutionary algorithms but the main ideas of these algorithms are different. The main goal of this paper is to discover how much these algorithms are usable and suitable to solve the problem of steel alloying.

Keywords: steel alloying, evolutionary algorithms, Differential evolution, SOMA

1 Introduction

1.1 Steel Alloying

It takes many steps to manufacture steel and every factory's step configuration may vary from the others. Commonly the process follows path similar to this: iron ore is smelted in iron smelters and distributed as a hot metal with temperature about 1400°C to converter or electric arc furnace to improve its quality for following treatments. Hot metal is mixed with scrap and slag formers, the heat is heated up approximately to 1650°C and first ferro-alloys are charged into the heat during tapping. This point is the first appliance of alloying process described below. Ladle with the heat is going to secondary metallurgy process, typically ladle furnace or vacuum treatment. This treatment part is exceptionally skipped, otherwise the steel quality is improved with accurate amount of alloys to achieve requested steel composition defined by steel grade and the temperature is modified for the last step. The final step of this process is casting on continuous casting machines to obtain final steel product like slabs, blooms or billets. More on this process can be found for example in book by Ahindra Ghosh and Amit Chatterjee [4].

Target of the whole process is primarily to manufacture steel of requested quality (this is commonly referred to as target or final steel grade), secondly to reduce cost of the materials and energy used in the process.

To manufacture steel of certain grade, with specific chemical, electrical and mechanical properties, it is crucial to be sure that the steel is composed of

correct elements in correct ratio's. Composition of the steel is modified in each step throughout the whole process with technique known as alloying. Alloying is process of continually charging the steel with certain amount's of alloying materials (compounds of chemical elements) to gradually obtain the requested steel quality

Commonly in practice this problem is solved with Dantzig's simplex algorithm, which is a linear programming solution that with given input parameters finds cheapest solution in given final steel's grade range. This however does not allow us to fine-tune optimal balance between precision of final solution and it's price. More on this matter can be found in book by Dantzig, George B. [5].

In this paper, we present how to overcome problems of the common solution with usage of evolution algorithms and fitness function suitable for solution of this problem is presented.

1.2 Evolution algorithms

The evolutionary algorithms have been successfully used to solve many practical and theoretical problems, see [1, 7, 8]. In this paper differential evolution and SOMA have been chosen.

Differential evolution (DE) is a population – based stochastic algorithm for global optimization. It was introduced by Ken Price and Rainer Storn and in this paper the original version is used for experiment design [3]. Although this algorithm is very simple and efficient, in [9] authors proved that it can not ensure the global convergence and they proposed two hybrids algorithms named QAISDE and GDISE to improve DE.

Unlike differential evolution, SOMA works at the principle of one population, which is changing in time. Application of this algorithm can be found for example in [11, 12]. In [13] the novel multiobjective SOMA has been introduced. In this paper the strategy All to one has been chosen [2]. For detailed description of algorithms see [14].

2 Experiment design

Our method is based on SOMA and DE using in steel alloying process. There were two main problems in experiment design - how to represent an individual in evolutionary algorithms and how to define its fitness function. From the view of steel alloying process, there were 14 compounds, which are added to the original steel composition. These compounds are consist of the elements. In result we needed the elements amount. In evolutionary algorithms the individuals consist of parameters. In this paper the parameters of individuals are represented by the compounds, which were added to the original steel composition. There were 14 compounds, so the individual's dimension (number of parameters) D has been 14. Each compound has been represented by one parameter in individual. In the end of the evolution the amount of the elements have been recomputed from the amount of compounds.

More on the fitness function design can be found in our paper [14].

2.1 Fitness Function Equation

If $Cmin_j < \left| \frac{M_i \cdot Mat_{ij} + S_j \cdot W}{\sum_{k=0}^{n_m} M_k + W} - C_j \right| < Cmax_j$ then fitness function is defined as follows (this is the case when given member represents valid solution that is in required composition range).

$$\sum_{i=0}^{n_m} \left[\left(\sum_{j=0}^{n_e} \left| \frac{M_i \cdot Mat_{ij} + S_j \cdot W}{\sum_{k=0}^{n_m} M_k + W} - C_j \right| \right) + P_i \cdot M_i \cdot PW \right] \tag{1}$$

Otherwise we multiply the result with additional penalization constant to equation to discard the member from solution.

Notation used in these equation is summarized in Table 1.

Table 1. The variables description from the Eqs. (1) and (??)

Variable	Description
n_m	Total number of available alloying materials (compounds) used.
n_E	Total number of chemical elements that materials consists of.
M	Individual being tested for it's fitness value. Bottom index specifies index of parameter whose value we want to obtain. Each parameter of member represents weight in kilograms (kg) of one material that would be charged to steel. Number of parameters of one member equals to number of all available alloying materials.
Mat	Material composition. First index specifies material, which's composition we are interested in. Second denotes index of element. Result is number in range $[0, 1]$ (1 means 100% of material is composed with that element, 0 that element is not contained in compound) that represents element's representation in given material (compound).
S	Original steel composition vector. Index denotes element whose representation we are interested in. Resulting number is again in range $[0, 1]$.
W	Original steel's weight in kg.
C	Desired composition vector. Index denotes element whose representation in steel we want to obtain. Resulting number is again in range $[0, 1]$.
$Cmin$	Desired minimal composition vector. Index denotes element whose representation in steel we want to obtain. Resulting number is again in range $[0, 1]$.
$Cmax$	Desired maximal composition vector. Index denotes element whose representation in steel we want to obtain. Resulting number is again in range $[0, 1]$.
P	Unit price of material for each kg.
PW	Constant defining importance of final solution's cost over it's precision.

More on the experiment design can be found in our publication [14].

3 Experimental Results

Our experimental results are summarized in Table 2. Highlighted rows represent runs that were unable to reach satisfying solution. First column of the table contains description of given settings along with it's parameters. Second column contains best achieved fitness value of all runs and next columns contain information about final computed solution of the steel.

From the results measured in our experiments we can see, that by using higher values then 0.5 for F in DE the evolution requires higher number of generations to successfully reach valid solution. With $F = 0.5$ we were able to reach satisfying solution even with only 50 members in whole population. Although some of the runs were unsuccessful and were unable to present valid solution, the runs that completed successfully yielded correct result reliably and we have not recorded single case when they would fail.

4 Conclusion

We have proposed alternative way to compute steel alloying recipe. Method using evolution algorithms has several advantages when compared to most commonly used Simplex method. It allows to incorporate final cost of used alloying materials into equation. Result of Simplex method is always cheapest solution in given range. Using our fitness function described in Section 2 we can control the importance of precise solution of final steel over it's price.

All runs presented in our experiments that were able to achieve less then approximately 155 fitness value provide satisfactory results and are fully prepared to be used in real life environment on real life cases.

We would like to extend our work and try out more evolution algorithms to compare their results. Mainly we would like to design larger experiment to compare performance and precision of larger set of evolution algorithms on this problem.

Acknowledgement This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and by the Development of human resources in research and development of latest soft computing methods and their application in practice project, reg. no. CZ.1.07/2.3.00/20.0072 funded by Operational Programme Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic.

References

1. Alguliev, R., Aligulizev R., Isazade, N.:DESAMC+DocSum: Differential evolution with self-adaptive mutation and crossover parameters for multi-document summarization, Knowledge - based systems, Pages: 21 - 38, DOI: 10.1016/j.knosys.2012.05.017, 2012

2. Onwubolu, GC., Babu, BV.: *New Optimization Techniques in Engineering*, Springer, ISBN 3-540-20167-X Springer - Verlag Berlin Heidelberg New York, 2004
3. Chakraborty, U., K.: *Advances in Differential Evolution*: Springer - Verlag Berlin Heidelberg, 2008, ISBN 978 - 3 - 540 - 68827 - 3
4. AHINDRA GHOSH, Amit Chatterjee. *Ironmaking and steelmaking: theory and practice*. New Delhi: Prentice-Hall of India, 2008. ISBN 978-812-0332-898.
5. DANTZIG, George B a Mukund Narain THAPA. *Linear programming*. New York: Springer, c1997-2003, 2 v. ISBN 03879861382-.
6. Zhang, Y., Gong, DW., Zhang, JH: Robot path planning in uncertain environment using multi-objective particle swarm optimization, *Neurocomputing*, Pages 172 - 185, DOI: 10.1016/j.neucom.2012.09.019, 2013
7. Chakaravarthy, GV, Marimuthu, S., Sait, AN.: Performance evaluation of proposed Differential Evolution and Particle Swarm Optimization algorithms for scheduling m-machine flow shops with lot streaming, *Journal of intelligent manufacturing*, Volume 24, Pages: 175 - 191, 10.1007/s10845-011-0552-2 , 2013
8. Subramanian, P. et al: PRISM: PRiority based SiMulated annealing for a closed loop supply chain network design problem, *Applied soft computing*, Volume 13, Pages: 1121 - 1135, DOI: 10.1016/j.asoc.2012.10.004, 2013
9. Chengfo, S.: Improved differential evolution algorithms, *Computer Science and Automation Engineering (CSAE)*, 2012 IEEE International Conference, 25 - 27 May 2012, Pages 142 - 145
10. Price, KV., Storn RM., Lampinen JA: *Differential Evolution A Practical Approach to Global Optimization*, Springer, 1997
11. Zelinka, I., Skanderova, L.: Investigation on Evolutionary Control and Optimization of Chemical Reactor, *Soft computing models in industrial and environmental applications*, *Advances in Intelligent Systems and Computing*, Volume 188, Pages: 469 -474, 2013
12. Pavlech, M.: Self-organizing Migration Algorithm on GPU with CUDA, *Soft computing models in industrial and environmental applications*, *Advances in Intelligent Systems and Computing*, Volume 188, Pages 173 - 182, 2013
13. Kadlec, P., Raida, Z.: A Novel Multi-Objective Self-Organizing Migrating Algorithm, *Radioengineering*, Volume 20, Pages 804 - 816, 2011
14. Holis, M. et al: Solving Steel Alloying Using Differential Evolution and SOMA, *CISIM* 2013

Comparison of Bio-inspired and Mathematical Optimization Methods on a Graph Drawing Problem

Lukas Rapant

Department of Applied Mathematics, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
lukas.rapant@vsb.cz

Abstract. We present comparison of bio-inspired and mathematical approach in a process of finding global minimum of an energetic function that is used in force directed layout algorithms. We have been faced with the issue of displaying large graphs. These graphs arise in the analysis of social networks. In order to find global minimum of an energetic function we employ two bio-inspired algorithms (Differential Evolution and Self-Organizing Migration Algorithm (SOMA)) and three mathematical optimization algorithm (Newton method, gradient descent and nonlinear conjugate gradient method). In this article basics of these algorithms, their results and comparison will be presented. We will also propose some forms of hybridization of these algorithms.

Keywords: mathematical optimization, bio-inspired optimization, force-directed layout

1 Introduction and problem statement

In this article, we will discuss and compare various ways to enhance force-directed layout algorithm. Force-directed algorithms calculate the layout of a graph using only information contained within the structure of the graph itself. However, they have two distinctive disadvantages: poor local minima, high running time. The problem of calculating graph layouts arose in the Indect project when we tried to analyze data from social networks. In this article we focus on the problem of finding the global minimum of an energetic function of these algorithms. To find the global minimum we will use five methods: differential evolution (DE) and SOMA, which belong to the bio-inspired optimization class, and Newton method (NM), Gradient descent (GD) and Nonlinear conjugate gradient method (NCG), which belong to mathematical optimization algorithms. We will also propose some ways to create hybrid algorithms, which take best of the both classes of optimization algorithms.

2 Theoretical background

2.1 Force-directed layout

Force-directed graph drawing algorithms are a class of algorithms for drawing graphs in an aesthetically pleasing way. Their purpose is to position the nodes of a graph in two-dimensional or three-dimensional space so that all the edges are of more or less equal length and there are as few crossing edges as possible, by assigning forces among the set of edges and the set of nodes, based on their relative positions, and then using these forces either to simulate the motion of the edges and nodes or to minimize their energy [1]. Energetic function describing such system can be defined as

$$\sum_{i=1}^{i \leq |V(G)|} \left(\sum_{j=1}^{j \leq |V(G)|} \sum_{i \neq j} \frac{1}{A(i,j)=0} \frac{1}{\sqrt{\sum_{k=1}^{k \leq 2} (x_{ik} - x_{jk})^2}} + \sum_{j=1}^{j \leq |V(G)|} \sum_{A(i,j) \neq 0} \ln \left(\sqrt{\sum_{k=1}^{k \leq 2} (x_{ik} - x_{jk})^2} \right) \right),$$

where where A is adjacency matrix of the graph, V(G) is set of all vertices of the graph and x is vector with Cartesian coordinates of all vertices. This function will be subject of our optimization algorithms.

2.2 Bio-inspired algorithms

In this part we will describe two bio-inspired algorithms: differential evolution and SOMA.

Differential evolution has proven to be an efficient method for optimizing real-valued functions. DE is trying to breed as good population of individuals as possible in loops called generations [2]. The very first step is defining constants affecting behavior of evolution algorithm. Those constants are crossover probability (CR) which influences probability of choosing a parameter of the mutated individual instead of the original one, mutation constant (F) which determines volume of mutation of the mutated individual, population size (NP) and finally trial specimen. Now we have a specimen and we are able to create initial population vector X. Whenever we have initial population prepared so called generation loop begins. In each generation cycle nested loop is executed. This loop is called evolution cycle. Each individual from current generation is bred to have better characteristic using mutations in evolution cycle. Algorithm finishes either if a solution is good enough (individual) or after defined number of generation loops had been executed. For our purposes the DE/rand/1 algorithm mutation is the best choice. The notation DE/rand/1 specifies that the vector v to be perturbed is randomly chosen and the perturbation consists of one weighted vector. Due to this mutation, new individuals are not affected by the temporary best individual from generation and space of possible solutions is searched through uniformly. More detailed description of Differential Evolution can be found in [2].

SOMA is a stochastic optimization algorithm that is using social behavior of cooperating individuals. Although the SOMA is not typical representative of evolution algorithm class because it is not creating new children while executing, it has very similar features as DE algorithm. SOMA is based on vector

operations and evolution-ary loops called migration loops. Whenever we specify control parameters and create a specimen we will randomly distribute all individuals of initial population over the search space. Each specimen computes its own fitness and the best individual becomes a Leader. Within each migration loop, all individuals travel certain distance of defined length (PathLength) towards the Leader making specified number of steps of certain length (Steps). Direction of these steps is however progressively deformed by perturbation vector (PRT). This vector forces specimen to explore larger area and avoid local minima. Each individual computes its fitness on each step position and at the end of the migration loop it returns back to the best path position. Once we find acceptable solution (diversity between the worst and the best specimens is lower than a defined value) the SOMA ends. More detailed description of SOMA can be found in [3].

2.3 Mathematical optimization methods

In this part three mathematical optimization algorithms will be described.

Newton method attempts to construct sequence (x_n) , that converges from some initial guess x_0 to x_* such that $f'(x_*) = 0$. If $f(x)$ is twice-differentiable, can be approximated by its second order Taylor polynomial and x_0 is chosen to be close enough to x_* , the sequence (x_n) defined by

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}, \quad n = 0, 1, \dots, \tag{1}$$

will converge towards the root of $f'(x)$. More detailed description of Newton method can be found in [4].

Main principal of Gradient descent is fact that if the function $f(x)$ is defined and differentiable in a neighborhood of a point x_0 , then $f(x)$ decreases fastest from x_0 in the direction of the negative gradient of $f(x)$ at x_0 , i.e. $-\nabla f(x_0)$. Based on this observation we can construct sequence x_0, x_1, \dots such that

$$x_{n+1} = x_n - \gamma_n \nabla f(x_n), \quad n = 0, 1, \dots, \tag{2}$$

where γ_n is small number representing length of a step in each iteration and is determined for example by line search. Providing that optimized functions satisfies certain conditions, such sequence converges to the local minimum. More detailed description of Gradient descent can be found in [4].

Nonlinear conjugate gradient method is generalization of conjugate gradient method for nonlinear optimization. Method is in some ways similar to the gradient descent method. First step is done exactly as in gradient descent method, but following steps are done in the conjugate direction

$$s_n = -\nabla x_n + \beta_n s_{n-1}, \tag{3}$$

where $-\nabla x_n$ is steepest descent direction and β_n is calculated either by Fletcher-Reeves, Polak-Ribiere or Hestenes-Stiefel formula. Sequence of (x_n) created by these conjugated steps converges towards local minimum. More detailed description of nonlinear conjugate gradient method can be found in [4].

3 Results

3.1 Comparison of algorithms

We have tested all five algorithms on a graph with 50 vertices. Best stopping criterion for mathematical optimization algorithms was found to be function value difference. This value has been set to 0.01 in all algorithms. Polak-Ribiere formula was chosen to be used for β computation in case of nonlinear conjugated gradient method. It also must be noted that while we have been able to obtain first derivative of objective function, second derivative is very difficult to obtain analytically. Therefore we are numerically approximating it in case of Newton method. In case of bio-inspired approaches, we have set the parameters in accordance with recommendations in [5]. These settings are written in Table 1.

Algorithm	Differential evolution	Algorithm	SOMA
Number of generations	100	Number of migrations	100
Population size	500	Population size	100
Crossbreeding threshold	0.95	Path length	3
Mutation constant	0.2	Step	0.2
		Perturbation	0.5

Table 1. Differential evolution and SOMA settings

Values of minimum found by tested methods with these settings are found in Figure 1. We have run 50 tests on the same graph with different starting points.

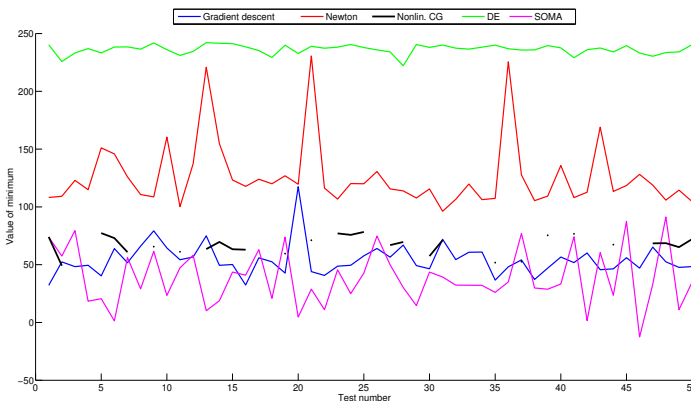


Fig. 1. Optimization algorithms results

It is evident from Figure 1 that the best results are given by gradient descent method and SOMA. Nonlinear conjugated gradients also gave us good results but they often had problems with convergence (i.e. they had not found minimum within 1000 iterations). This is represented by blank parts of line in in Figure 1. It is also fair to note that, during testing, gradient descent method also had problems with convergence though they were quite rare. Newton method gave use much higher values and differential evolution even higher. However, in our previous work we have proven that differential evolution can be tailored to find much better minima, but it requires extensive parameter optimization.

Value of the minimum is not the only measure of quality in our application. Evaluation of our objective functions is quite time consuming and even more so with bigger graph as its computational complexity is n^2 . Therefore we must also measure quality of the algorithm by number of objective function evaluation. In this case, best algorithm is nonlinear conjugated gradients and it is followed by gradient descent method (on our testing graph NCG had 278 function enumerations and GD had 537 enumerations in average). It must be noted that nonlinear conjugated gradients performs even better as it only evaluates derivative of the objective function, which is evidently much less time consuming, while gradient descent evaluates both objective function and its derivative. Bio-inspired techniques and Newton method perform much worse in this measure as all three methods have about 1200 function evaluations. Newton method is again better then DE and SOMA as it evaluates derivative of the objective function.

When both of these parameters are taken into consideration, the best methods seem to be SOMA and gradient descent method. While SOMA offers usually better solution and more reliably, gradient descent is much faster. While nonlinear conjugated gradients offer good solutions in the best time, this method is too unreliable in our case. Newton method and differential evolution can be ruled out as slow and, in case of DE, too parameter sensitive.

3.2 Hybrid methods

We have also done some preliminary testing with hybrid algorithms, i.e. algorithms containing both bio-inspired and mathematical approach. Based on results from our previous research, we have chosen SOMA algorithm and gradient descent algorithm. We have tried two approaches. In the first one (named SOMAIn), we have tried to optimize positions of all members of population after each migration by gradient descent. In the second one (named SOMAOut), we have done this optimization after SOMA algorithm has finished. These approaches have been tested on the random graph with 50 vertices and compared to mundane SOMA. We have set the algorithms in such way, that each variant has approximately same number of function evaluations. Average results after 50 tests are following: SOMA has found average minimum of 46.35, SOMAIn average minimum of 59.56 and SOMAOut average minimum of 43.78. It is evident, that the best approach in our case is SOMAOut approach as even more migrations and members of population cannot provide improvement as one gradient descent algorithm applied on SOMA result. SOMAIn finds better minimum

than SOMA, but when we provide SOMA additional migrations and population members instead of gradient descent algorithm runs, we obtain better result.

4 Conclusion

In this article we have tested five algorithms for optimizing of function underlying force-directed graph layout. We have shown that the best algorithms are SOMA and gradient descent algorithm. Main question is how would all presented algorithms perform in case of very large graph. Prerequisite for these test is parallelization of these methods and this will be done in near future. We have also shown some experiments with hybridization of these algorithms and proven that this way can be beneficial. This approach will also be subject of future research.

5 Publications

Parts of this research were published in the following publications:

- Patrik Dubec, Jan Plucar, Luk Rapant, Use of Bio-Inspired Algorithms to Find Global Minimum in Force Directed Layout Algorithms, Proceedings of MCSS 2013, Communications in Computer and Information Science 368, pp. 194-203, Cracow, 2013, ISSN 1865-0929.
- Patrik Dubec, Jan Plucar, Luk Rapant, Case Study of Evolutionary Process Visualization Using Complex Networks, Proceedings of Nostradamus 2013, Advances in Intelligent Systems and Computing 210, pp. 125-135, Ostrava, 2013, ISSN 2197-5357.

References

1. Kobourov S. G., Force-directed drawing algorithms. In R. Tamassia, editor, Handbook of Graph Drawing and Visualization, CRC Press, 2012.
2. Price K., An Introduction to Differential Evolution, in New Ideas in Optimization, D. Corne, M. Dorigo and F. Glover, Eds., pp. 79108, McGraw-Hill, London, UK, 1999.
3. Zelinka I., SOMA Self Organizing Migrating Algorithm, in New Optimization Techniques in Engineering, Eds.: Babu B. V. and Onwubolu G. (Springer-Verlag, New York), pp. 167-218, ISBN 3-540-20167X, 2004.
4. Snyman J., Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms (Applied Optimization), Springer-Verlag, New York, 2005, ISBN 978-0387298245.
5. Price K., and Storn R., Differential Evolution A simple evolutionary strategy for fast optimization, Dr. Dobb's Journal 264, Miller Freeman, Inc., San Francisco, 1997.

Architectures and Features of Modern Honeybots

Daniel Stríbný and Zbyněk Složil

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
{daniel.stribny, zbynek.slozil.st}@vsb.cz

Abstract. A honeypot is a trap set to detect, deflect or in some manner counteract attempts at unauthorized use of information systems. It must be taken into consideration that honeypot is designed to be compromised, so attacker could take advantage of this fact. These data must be accordingly processed, so various data mining techniques are often deployed. We propose data visualization and processing by a specialized software - Orange and others. These can be also done by an dedicated application, that is build in to the honeynet framework as we also propose in design of our honeynet, which is currently being built.

Key words: honeypot, honeynet, data mining, visualization

1 Introduction

Hackers exploit more tricky and obscure methods. Both random and targeted attacks are standard practice on a public network. Such attacks have many forms and various sizes. Against them are used various firewalls, which can block defined IP addresses and ports, or are used Intrusion Detection System (IDS), that can recognize more complex attacks by their signatures. A disadvantage of these traditional protections is that they can not respond quickly to new types of attacks. In many cases, it leads to adding some kind of a honeypot to the protective system. This honeypot is intended to be attacked and compromised to gain more information about the attacker and his attack techniques. In Figure 1 [1] can be seen, the number of attacks on specific services (Active Directory, HTTP, ...), captured on honeypots from CZ.NIC [2].

1.1 Honeypot

A honeypot is a trap set to detect, deflect or in some manner counteract attempts at unauthorized use of information systems. Usually contains decoy data, computers or various network devices that appears to be part of a network, but is actually isolated and monitored, and seems to have some value to potential attackers.

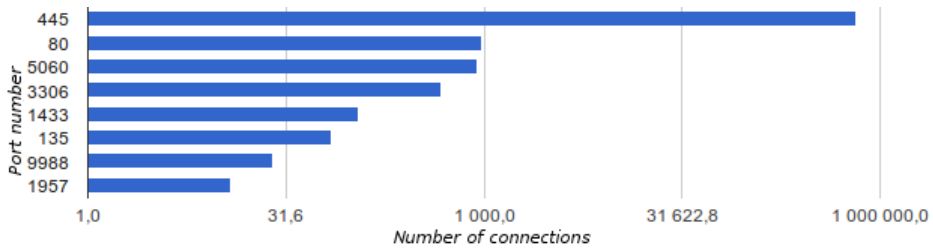


Fig. 1. Target ports on NIC honeynet (November 2012 - January 2013)

The honeypot hasn't a production value or an authorized activity, but it is a highly flexible security tool. It doesn't fix any problem but instead it has important use in intrusion prevention, detection and information gathering. It must be taken into consideration that the honeypot is designed to be compromised, so attacker could take advantage of this fact. [6].

Honeypots are decoys set up for the purpose of monitoring and logging the activities of entities that probe, attack or compromise them. Activities on honeypots can be always considered as suspicious, as there is no point for normal users to interact with them. Honeypots come in many variants; examples include dummy items in a database, low-interaction network components like pre-configured traffic sinks, or full-interaction hosts with real operating systems and services. Honeypots reduce false positives by capturing small data sets of high value, capture unknown attacks, and work in various environments. [9]

Types

Based on deployment:

Production honeypots are placed inside a production network with other production servers by an organization to improve their overall state of security. Normally, they are low-interaction honeypots, which are easier to deploy. They give less information about the attacks or attackers than research honeypots do.

Research honeypots are complex to deploy and maintain, capture extensive information, and are used primarily by research, military or government organizations. They are used to research the threats organizations face and to learn how to better protect against threats. [8]

Based on design criteria:

Pure honeypots are full-fledged production systems. The activities of an attacker are monitored using a casual tap that has been installed on the honeypot's link to a network.

High-interaction honeypots imitate activities of the real systems, that host a variety of services, and therefore an attacker may be allowed a lot of services to waste his time. In general, high interaction honeypots provide more

security by being difficult to detect, but they are highly expensive to maintain.

Low-interaction honeybots simulate only the services frequently requested by attackers. Since they consume relatively few resources, multiple virtual machines can easily be hosted on one physical system, the virtual systems have a short response time, and less code is required, reducing a complexity of a security of virtual systems.

A honeynet is a network of two or more honeybots. This network is assembled from both high and low interaction honeybots that simulates a production network and configured such that all activity is monitored, recorded and in a degree, discreetly regulated. It is used for monitoring a larger or more diverse network in which one honeybot may not be sufficient.

Honeybots often produce large quantity of logs from captured attacks on their output. These data must be accordingly processed, so various data mining techniques are often deployed.

1.2 Data mining

In order to improve honeybot performance, various data mining techniques being often applied to log analysis. Data mining is one of the steps in the Knowledge Discovery in Databases (KDD) process. A simple description of KDD is a general discovery process of useful knowledge (previously unknown) from large databases.

Data from honeybots and honeynets are ideal input for various data mining methods. These methods are able to reduce the amount of logs which will require human attention. Most common examples of data mining algorithms are k-nearest neighbor and decision tree.

The k-nearest neighbor hypothesis is that elements of the same class will be close to each other in the attribute space. The KNN works as follows: for each instance of data of unknown class it calculates the attribute space distance between it and all instances with known classes, selecting as the class for the unknown one the most present class in the K nearest instances.

A decision tree is a kind of machine learning algorithm known as successive partitioning algorithm. Its goal is to build a model that predicts the value of a target variable based on several input variables.

These algorithms work with the partitioning of the original data set in successively more homogeneous subgroups, which are partitioned in nodes until the desired detail level is accomplished [4][5].

2 Research

We propose usage of suggested data mining methods to classify honeybot data if that data originate from a honeynet and are not preprocessed automatically. In Figure 2 is a visualization of data obtained from honeybots of Microsoft Malware

Protection Center[7] processed by a Orange data mining tool[3]. Therefore is not so surprising that most attacked ports are 445 - MS SMB and 135 - MS DCE/RPC Locator service. Biggest drawback of these publicly available data is anonymization of source and destinations IP addresses.

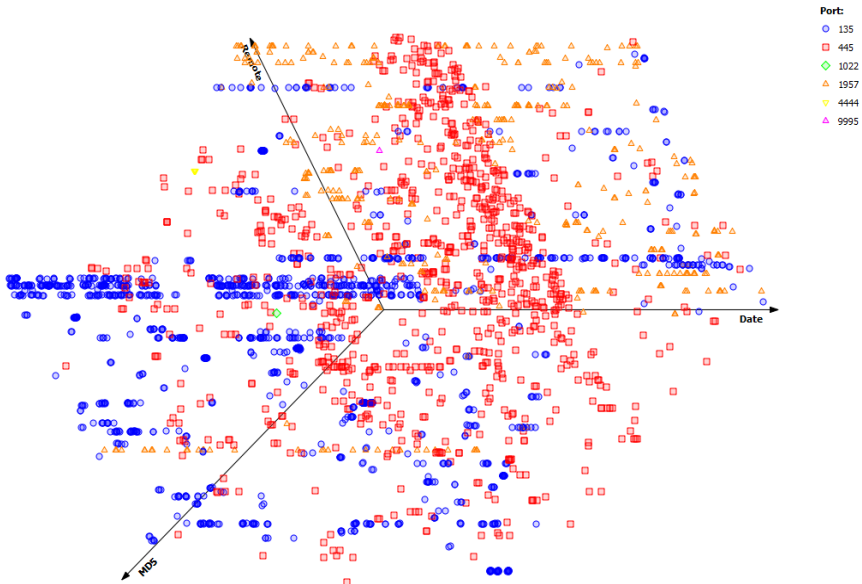


Fig. 2. Visualization of data sample from honeypot

We are also in process of building a local honeynet. It will be composed from both high and low interaction honeypots. It will run some simulated Cisco routers and most common web services. We also developing some kind of a WiFi honeypot to gather mobile users attacks on WiFi hotspots. These Wifi honeypots are secured with an WEP algorithm or have an open authentication. Their value is not as high as by traditional honeypots but reflects todays trend of a heavy smartphone and tablet usage. All data captured on this honeynet will be visualized and available for researchers.

Schematic of our honeynet is in Figure 3.

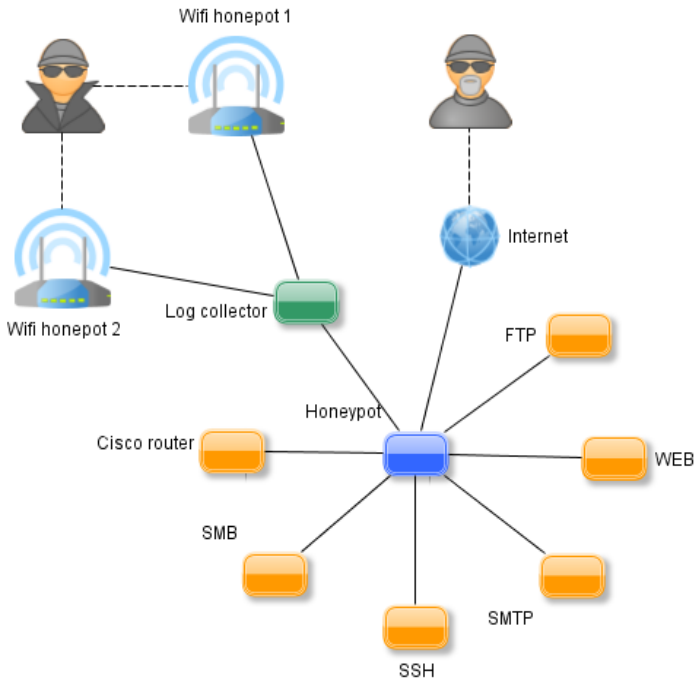


Fig. 3. Design schematic of our honeynet

3 Conclusion

In this paper, we introduced various types of honeypots. We propose for usage data mining techniques to gather important data from honeypot logs. We also propose a design of the honeynet, which is currently being built and which is expected to obtain various data for our further research, as these data are very difficult to obtain.

References

1. CZ.NIC, z. s. p. o. .blog - Honeynet: utoky na prelomu roku, <https://blog.nic.cz/2013/02/19/honeynet-utoky-na-prelomu-roku/>, 2013.
2. CZ.NIC, z. s. p. o. CZ.NIC, <http://www.nic.cz/>, 2013.
3. Faculty of Computer and Information Science, University of Ljubljana, Slovenia. Orange, Data Mining Fruitful & Fun, <http://orange.biolab.si/>, 2013.
4. U. Fayyad, G. Piatetsky-shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, 17:37–54, 1996.
5. A. Grgio, R. Santos, and A. Montes. Evaluation of data mining techniques for suspicious network activity classification using honeypots data. pages 657006–657006–10, 2007.
6. Lance Spitzner. The Value of Honeypots, Part One: Definitions and Values of Honeypots, <http://www.symantec.com/connect/articles/value-honeypots-part-one-definitions-and-values-honeypots/>, 2001.
7. Microsoft Malware Protection Center. Carnivore news - berlin raw data, <tp://ftp.carnivore.it/projects/dionaea/rawdata/berlin-20091207-logsql.sqlite.bz2>, 2009.
8. L. Spitzner. *Honeypots: tracking hackers*, volume 1. Addison-Wesley Reading, 2003.
9. U. Thakar, S. Varma, and A. K. Ramani. Honeyanalyzer : Analysis and extraction of intrusion detection patterns & signatures using honeypot. In *In Proceedings of the IEEE Infocom 2005, 2005*. [46] Urjita Thakar, Sudarshan.

Growing Neural Gas – A Parallel Approach

Lukáš Vojáček

Department of Computer Science, FEECS,
VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava – Poruba
lukas.vojacek@vsb.cz

Abstract. The paper deals with the high dimensional data clustering problem. One possible way to cluster this kind of data is based on Artificial Neural Networks (ANN) such as SOM or Growing Neural Gas (GNG). The learning phase of the ANN, which is time-consuming especially for large high-dimensional datasets, is the main drawback of this approach to data clustering. The parallel modification, Growing Neural Gas, and its implementation on the HPC cluster is presented in the paper. Some experimental results are also presented.

1 Introduction

Recently, the issue of high-dimensional data clustering has arisen together with the development of information and communication technologies which support growing opportunities to process large data collections. High dimensional data collections are commonly available in areas like medicine, biology, information retrieval, web analysis, social network analysis, image processing, financial transaction analysis and many others.

Two main challenges should be solved to process high-dimensional data collections. One of the problems is the fast growth of computational complexity with respect to growing data dimensionality. The second one is specific similarity measurement in a high-dimensional space. As presented in [1], Beyer et al. for any point in a high-dimensional space the expected distance, computed by Euclidean measure to the closest and to the farthest point, shrinks with growing dimensionality. These two reasons reduce the effectiveness of clustering algorithms on the above-mentioned high-dimensional data collections in many real applications.

The authors propose an effective data clustering algorithm which is based on *Growing Neural Gas* (GNG) [5]. The computational complexity is resolved by the parallel implementation of GNG. Some technical problems have to be resolved in order to effectively train such kind of neural network using an *High Performance Computing* (HPC) cluster with MPI. The traditional serial approach to training GNG is also considered in the paper. The serial learning GNG algorithm is used for benchmarking the parallel version of GNG. In other words, parallel GNG has to produce the same network and should be an order of magnitude faster in the ideal case.

2 Artificial Neural Network

2.1 Growing Neural Gas

The principle of this neural network is an undirected graph which need not be continuous. Generally, there are no restrictions on the topology. The graph is generated and continuously updated by competitive Hebbian Learning [4, 6]. According to the pre-set conditions, new neurons are automatically added and connections between neurons are subject to time and can be removed. GNG can be used for vector quantization by finding the code-vectors in clusters [3], biologically influenced [7], image compression, disease diagnosis.

GNG works by modifying the graph, where the operations are the addition and removal of neurons and edges between neurons.

To understand the functioning of GNG, it is necessary to define the algorithm. The algorithm described by Algorithm 1 is based on the original algorithm [2] [3], but it is modified for better continuity in the SOM algorithm. The description of the algorithm has been divided for convenience into two parts. In the first part of Algorithm 1 the overall functionality is described.

Remark The notation used in the paper is briefly listed in Table 1.

Table 1. Notation used in the paper

Symbol	Description
n	Dimension of input vectors, number of input neurons, dimension of weight vectors in GNG output layer neurons
N	Current number of neurons in GNG output layer
N_{max}	Maximum allowed number of neurons in GNG output layer
E	set of all edges in GNG
G	undirected graph describing topology of GNG, $G(\{N_1, \dots, N_N\}, E)$
T	Number of epochs
t	Current epoch, $t = 1, 2, \dots, T$
$\mathbf{w}_k(t)$	Weight vector of neuron N_k , $k = 1, 2, \dots, N$ $\mathbf{w}_k(t) \in \mathbb{R}^n$, $\mathbf{w}_k(t) = (w_{1k}, w_{2k}, \dots, w_{nk})$
e_i	Local error of output neuron N_i , $i = 1, 2, \dots, N$
α	Error e_i reduction factor
β	Neuron error reduction factor
γ	Interval of input patterns to add a new neuron
a_{max}	Maximum edges age
a_{ij}	Age of edge e_{ij}

3 Parallelization

Parallelization focuses on the equally distribution of neurons, where new neurons are allocated to the process with the lowest number of neurons. The advantage

Algorithm 1 Growing Neural Gas algorithm

1. Initialization of network. Two neurons N_1 and N_2 are created, $E = \{e_{12}\}$. Weight vectors $w_1(t)$ and $w_2(t)$ are initialized to random values $w_{kj}(t) \in [0, 1]$.
 2. Select arbitrary unused input data vector.
 3. Perform the one learning iteration.
 4. Reduce error value e_i for all neurons N_i using factor β .
 5. Returns to step 2, until all input data vector have been used.
 6. If $t < T$ return to step 2.
-

of this distribution is constant workload processes. The disadvantage is increase communication between processes.

After analysing the GNG learning algorithm we identified the one most processor time-consuming area. This part was selected as a candidate for the possible parallelization. The selected area is:

Finding BMU – this part of GNG learning can be significantly accelerated by dividing the GNG output layer into smaller pieces. Each piece is then assigned to an individual computation process. The calculation of the Euclidean distance among the individual input vector and all the weight vectors to find BMU in a given part of the GNG output layer is the crucial point of this part of GNG learning. Each process finds its own, partial, BMU in its part of the GNG output layer. Each partial BMU is then compared with other BMUs obtained by other processes. Information about the BMU of the whole network is then transmitted to all the processes to perform the updates of the BMU neighbourhood.

3.1 Description of Proposed Approach

This subsection includes a detailed description of our approach. Initially, each process reads the training input vectors and on the first process are create two neurons. During the calculation, new neurons are equally distributed on the processes. We do not store the complete GNG graph, we only store parts of it in the appropriate computational nodes.

In the main computational phase, one BMU and a second BMU are founded for each input vector. Thus, each processor needs to compute its local BMU and second BMU within its neurons, after which, each local BMU and second BMU (and their position in the network) are shifted onto one process using the MPI function *GatherFlattened* to determine the global BMU and second BMU. It is possible to use another MPI functions as well, which can provide this selection at one time, but after testing we have found that the experiments took much more time than our presented approach. A global winning BMU and second BMU are then distributed using the MPI function *Broadcast* on all the processes. Now if there is a edge between the first BMU and the second BMU then age is set to zero otherwise creates edge between this two neurons. Next, the neighbourhood of the BMU in each process is known and, consequently, the

weights of the neurons matching are actualized. If the condition is met for adding a new neuron, the process with the lowest number of neurons add a new neuron. This procedure is repeated until all the input training vectors are exhausted (until we have finished one epoch).

4 Experiments

4.1 Experimental Datasets and Hardware

Two datasets were used in the experiments. The first dataset was commonly used in Information Retrieval – *Medlars*. The second one was the test data for the elementary benchmark for clustering algorithms[8].

4.2 First Part of The Experiment

The first part of the experiment was oriented towards a comparison of the standard GNG algorithm and parallel approach to this GNG learning algorithm. The Medlars dataset was used for the experiment. A parallel version of the learning algorithm was run using 2, 8, 16, 24 and 32 MPI processes. The records with an asterisk (*) represents the results for only one process i.e. this is the original serial learning algorithm and there is no network communication.

GNG parameters are the same for all experiments and are as follows $\gamma = 200$, $e_w = 0.05$, $e_n = 0.006$, $\alpha = 0.5$, $\beta = 0.0005$, $a_{max} = 88$, $M = 1000$, $\delta = 100$. The achieved computing time is presented in Table 2.

Table 2. Computing Time with Respect to Number of Cores, Standard GNG Algorithm, Dataset Medlars

Cores	Computing Time [hh:mm:ss]
1*	00:35:41
2	00:17:50
8	00:04:47
12	00:03:56
16	00:03:09
24	00:02:45
32	00:02:32

As we can see from Table 2, the computing time depends on the number of used cores as well. With a growing number of processors, the computation effectiveness increases, and the computational time is sufficiently reduced.

4.3 Second Part of The Experiment

The second part of the experiments was oriented towards comparing the results obtained by the parallel and standard GNG algorithm. The Clustering dataset

was used for the experiment. The parallel version of the learning algorithm was run using 16 MPI processes.

GNG parameters are similar to the previous experiment. There are two changes $M = 500$ and $\delta = 25$.

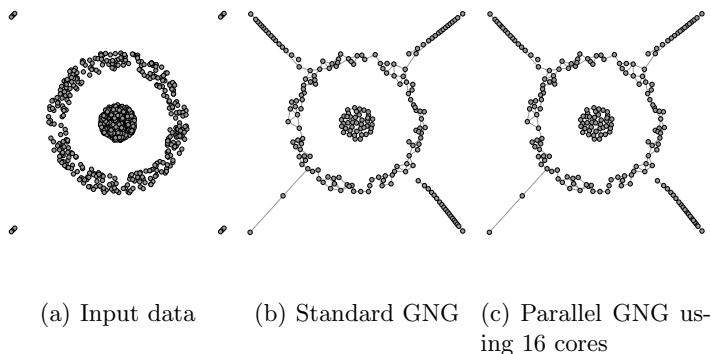


Fig. 1. Results of dataset *Target*

In Figure 1(a) there is a layout view input data, which is used for training GNG. Outputs of standard GNG algorithm are in Figure 1(b), which is the same as in the parallel version. i.e. both versions produce the same network.

5 Conclusion

In this paper the parallel implementation of the GNG neural network algorithm is presented. The achieved speed-up was very good and the results from the standard and parallel version of GNG are same. So we can say that the operation of the parallel version is correct. However, the effectiveness of a parallel solution is dependent on the division of the output layer. An improper division may cause the communication between processes to be very time consuming.

In future work we intend to focus on the sparse data, use combinations of neural networks for improved result and improved acceleration.

6 Publications

1. J. Martinovič, K. Slaninová, L. Vojáček, P. Dráždilová, J. Dvorský, I. Vondrák. Effective clustering algorithm for high-dimensional sparse data based on SOM. In *Neural Network World*, volume 23, pages 131-147, 2013, ISSN 1210-0552, IF 0.646
2. L. Vojáček, and J. Dvorský. Growing Neural Gas – A Parallel Approach. In *IEEE Proceedings of International Conference on Computer Information Systems and Industrial Management Applications CISIM 2013*, 2013

Acknowledgement

This article is based on paper published on CISIM 2013.

References

1. K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft. When is "nearest neighbor" meaningful? In *Database Theory '99*, volume 1540, pages 217–235, 1999.
2. B. Fritzke. A growing neural gas network learns topologies. In *Advances in Neural Information Processing Systems 7*, pages 625–632. MIT Press, 1995.
3. J. Holmström. Growing neural gas experiments with gng, gng with utility and supervised gng. Master's thesis, Uppsala University, 2002-08-30.
4. T. Martinetz. Competitive hebbian learning rule forms perfectly topology preserving maps. In S. Gielen and B. Kappen, editors, *ICANN 93*, pages 427–434. Springer London, 1993.
5. T. Martinetz and K. Schulten. A "neural-gas" network learns topologies. *Artificial Neural Networks*, 1:397–402, 1991.
6. Y. Prudent and A. Ennaji. An incremental growing neural gas learns topologies. In *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, volume 2, pages 1211 – 1216 vol. 2, july-4 aug. 2005.
7. I. Sledge and J. Keller. Growing neural gas for temporal clustering. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, 2008.
8. A. Ultsch. Clustering with SOM: U*C,. *Proc. Workshop on Self-Organizing Maps, Paris, France,*, pages 75–82, 2005.

Author Index

- Babskova, Alisa, 511
Bašniar, Václav, 29
Bednář, Pavel, 35
Berek, Petr, 453
Bilik, Petr, 266
Blaheta, Radim, 345
Bobrov, Pavel, 459
Bocheza, Jiří, 41, 409
Borski, Martin, 1
Buriánek, Tomáš, 501
- Carbol, Zbyněk, 213
Cerny, Martin, 294
Cubik, Jakub, 404
- Dohnálek, Pavel, 465
Dubec, Patrik, 517
Dvorsky, Jiri, 523
- Đurica, Maroš, 147
- Fajkus, Marcel, 404
Friedrich, Jiří, 237
Frolov, Alexander, 459
Fusek, Radovan, 46, 88, 100
- Ganiyev, Artem, 409
Gasz, Rafał, 112
Grunt, Ondřej, 327
- Hapla, Václav, 333
Hasal, Martin, 339
Haslinger, Jaroslav, 345
Havel, Ales, 24
Hejduk, Stanislav, 41, 409
Helštýnová, Barbara, 153
Hercík, Radim, 261
Holis, Michal, 523
Holuša, Michael, 52
Horinek, Marek, 266
Hořejší, Jaroslav, 471
Houdek, Vít, 160
Hradílek, Zdeněk, 190
Hrtus, Rostislav, 345
Hrvolova, Barbora, 311
Húsek, Dušan, 459
- Ivánek, Lubomír, 434
- Jahan, Ibrahim Salem, 489
Janša, Jiří, 166
Janurová, Kateřina, 351
Jaros, Jakub, 415
Jeżowicz, Tomáš, 477
Jirka, Jakub, 58
- Kalvar, Daniel, 12
Kašpar, Adam, 471
Kepak, Stanislav, 404
Klimowski, Marek, 118
Konečný, Jaromír, 272
Košinár, Michael A., 64
Krček, Jiří, 357
Krohova, Jana, 311
Krutil, Jonáš, 483
Kučera, Radek, 369
Kudělka, Miloš, 477, 483
- Látal, Jan, 441
Liner, Andrej, 415
Lukáš, Petr, 70
- Macek, Jakub, 76
Mañas, Jakub, 277
Marcinčin, Oliver, 172
Meca, Ondřej, 82
Medvec, Zdeněk, 172
Merta, Michal, 363
Mikulec, Martin, 421
Motyčková, Kristina, 369
Mozdřeň, Karel, 46, 88, 100
Mynarek, Piotr, 124
- Niedźwiecki, Marcin, 128
Novák, Tomáš, 153
Nowaková, Jana, 282
- Opluštěl, Josef, 12
Otevřel, Vít, 288
- Papes, Martin, 415
Partila, Pavol, 428
Peter, Lukas, 294
Pětvaldský, Petr, 300
Placek, Martin, 523
Platoš, Jan, 477, 495
Plucar, Jan, 517
Pokorný, Miroslav, 282
Prílepok, Michal, 453, 489

Przybyła, Tomasz, 134
Pyka, Tomasz, 138

Rapant, Lukas, 529
Reguła, Michal, 178
Robenek, Daniel, 495
Ronovský, Aleš, 375, 392
Rozehnal, Petr, 184
Rozhon, Jan, 421
Rusnok, Stanislav, 243

Řezáč, Filip, 421

Safarik, Jakub, 428
Skalny, Pavel, 381
Skanderova, Lenka, 523
Slabý, Roman, 306
Slanina, Zdeněk, 288
Slíva, Ladislav, 249
Složil, Zbyněk, 535
Smoček, Martin, 190
Smondrk, Maros, 311
Snášel, Václav, 453, 477, 483, 489, 495
Sobota, Pavel, 243
Sojka, Eduard, 46, 88
Sokanský, Karel, 153, 213
Stachoň, Martin, 386
Stankuš, Martin, 306
Stolfa, Jakub, 94
Stříbný, Daniel, 535
Stuchlý, Jindřich, 196
Svoboda, Pavel, 243
Szabó, Dominik, 201

Špička, Jiří, 6

Šprlák, Roman, 12
Šrámek, Jan, 207
Šulák, Radek, 18
Šumpich, Jan, 213
Šurkala, Milan, 46, 88, 100
Šurkovský, Martin, 82
Švecová, Lucie, 317

Thao, Phan Thi Thanh, 434
Tintěra, Jaroslav, 459
Tomala, Karel, 428
Tvrdon, Martin, 24

Uher, Marián, 219
Uchoč, Miroslav, 12
Urbasek, Jiri, 106

Vaculík, Jan, 225
Vasatova, Alena, 392
Verner, Tomáš, 255
Vích, Lukáš, 76
Vitásek, Jan, 441
Vlček, Jaroslav, 357
Vojáček, Lukáš, 541
Vojčínák, Petr, 321
Vozda, Michal, 311
Vozňák, Miroslav, 421
Vramba, Jakub, 231
Vychodil, Jiri, 421

Witas, Karel, 409

Zaorálek, Lukáš, 501
Zapletal, Jan, 398
Zelinka, Ivan, 523

Editors: Michal Krátký, Jiří Dvorský, Pavel Moravec
Dept. of Computer Science
VŠB – Technical University of Ostrava

Title: WOFEX 2013

Place, year, edition: Ostrava, 2013, 1st

Page count: 565

Edit: VŠB – Technical University of Ostrava,
17. listopadu 15, 708 33 Ostrava-Poruba
Czech Republic

Impression: 200

Not for sale.

ISBN 978-80-248-3073-5